AD/COM

ADCOM, INC.
808 Memorial Drive
Cambridge 39, Mass.
UN 8-7386

ADVANCED COMMUNICATION RESEARCH AND DEVELOPMENT

AD/COM

First Quarterly Report

ADVANCED THRESHOLD REDUCTION TECHNIQUES STUDY

1 July 1964 - 30 September 1964

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland

under

Contract No. NAS 5-9011

by

ADCOM, Inc.
808 Memorial Drive
Cambridge, Massachusetts

AD/COM

First Quarterly Report

ADVANCED THRESHOLD REDUCTION TECHNIQUES STUDY

1 July 1964 - 30 September 1964

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland

under

Contract No. NAS 5-9011

Approved by Elie J. Baghdady
Elie J. Baghdady
Technical Director

by

ADCOM, Inc.
808 Memorial Drive
Cambridge, Massachusetts

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF ILLUSTRATIONS

I. GENERAL INTRODUCTION

This document constitutes the First Quarterly Report submitted by ADCOM, Inc. to the NASA Goddard Space Flight Center under Contract No. NAS 5-9011. The work covered here was performed at ADCOM, Inc. between 1 July 1964 and 30 September 1964.

The primary objective of this program is to demonstrate the feasibility of reducing the noise threshold for demodulating FM signals at least 3db below the threshold experienced with a second-order phase-locked loop demodulator whose design is optimized for best threshold performance in demodulating a given signal.

The technical plan of the program calls for a critical assessment of the current state of knowledge of the most promising threshold reduction approaches, taking into full account the available technical literature. The object of this assessment is to identify the various mechanisms that cause the threshold effect and limit the ultimate threshold reduction accomplishment with each of the various promising demodulation techniques. Inasmuch as the literature on this subject abounds in widely quoted analyses and results of a controversial nature, critical reviews of some of these publications are in order.

In the light of the results of the above critical examination of the various approaches, a few of the most promising approaches are to be singled out for implementation and thorough testing.

Section II of this report presents a synopsis of results reported in the main body of the report.

Section III presents a critical treatment of certain fundamental aspects of conventional demodulators, phase-locked loop, frequency compressive feedback, and band-dividing demodulators that are essential in the analysis of performance, evaluation of improvement potentialities, and comparison of ultimate performance limitations of each of these techniques.

Section IV presents a critical review and evaluation of the validity and significance of certain published analyses, and an extension of the results of one of the reviewed papers.

In Section V, analyses are presented of higher-order phase-locked loops to determine their ultimate optimized performance capabilities and requirements.

In Section VI, some aspects of experimental test and evaluation are discussed and an automated procedure is described.

The work plans for the next interval are outlined briefly in Section VII.

II. SYNOPSIS OF REPORT

The present chapter summarizes the main investigations reported in the sections that follow.

Chapter III characterizes the basic models of the PLL and FCF demodulators. The PLL is first characterized as a nonlinear, randomly time-variant system and the necessary assumptions involved in the conventional linear time-invariant model are indicated. An analogous analysis is then effected for the FCF demodulator. The absolute lower limit on closed-loop noise bandwidth is established for both systems and threshold reduction capabilities are considered on this basis. Band-dividing demodulation is finally considered and the limitations of different signal processing techniques are established. The case of CW signals is shown to yield a bandwidth comparable to that of a second-order PLL demodulator for each of the filters in the band-dividing bank so that the SNR performance of the Akima system will not show the desirable improvements. The case of sampling and quantizing is also eliminated because of the large number of filters required leaving sample-and-hold processing as the other possibility to be considered.

Chapter IV evaluates certain published analyses of interest. A detailed discussion of some of the limitations of Develet's papers is first presented. The work of Sanneman and Rowbotham is then evaluated and some basic results are clarified and extended.

Chapter V evaluates the performance of a second-order PLL demodulator and analyzes the possible threshold improvements attainable with higher order PLL's. The second-order loop performance is first optimized for a sinusoidal modulation and the above-threshold operational requirement is established. An analogous approach is used for the third order loop and the improvement capabilities are identified and found to be limited from feasibility considerations. The optimum realizable PLL demodulator is then determined for the cases of rectangular and an RC PM spectra. The threshold characteristics of these receivers are established and possible approximations to the optimum transfer function are finally presented and evaluated.

Chapter VI is concerned with the experimental program. The previous ADCOM work of SNR measurements is summarized and a threshold test system for the present contract is suggested. The necessary modifications that will yield an automatic, rather than manual, threshold curve plotter are included.

III. BASIC MODELS OF EXPONENT DEMODULATORS

In the analysis and comparison of threshold reduction techniques, it is of the essence that the basic models and assumptions used correspond to the physical conditions encountered in the real world. In the last few years, a number of attempts have been made at determining and comparing the ultimate performance of various exponent demodulation techniques. Unfortunately, many of these publications have introduced and perpetuated a host of now widespread misconceptions about the basic theoretical models and the fundamental considerations that define the performance characteristics, limitations, and comparative merits of these techniques.

Inasmuch as the major objective of this program is to demonstrate the feasibility of reducing the noise threshold of exponent demodulation at least three db below the threshold experienced with a second-order phase-locked loop, it is important that we re-examine the basic models and considerations that are essential for predicting the ultimate threshold reduction performance of modulators. This re-examination is essential for dispelling certain published analyses from the picture lest their results be considered to define true limits inhibiting further consideration of otherwise promising approaches. On the positive side, the re-examination brings into sharp focus the basic criteria and causes of the noise thresholds of various demodulators. In this way improved insight into the avenues for improving performance should result.

3.1 The Conventional FM Demodulator

The conventional limiter-discriminator demodulation technique provides a convenient standard of comparison for other demodulators. The performance characteristics of this type of demodulator have been the subject of many analyses and discussions, and hence will not be taken up here. However, one point is considered to be of sufficient importance here to deserve presentation.

The point of interest concerns the common assumption about the dependence of the threshold of conventional FM demodulation upon the deviation ratio of the anticipated FM signal. Thus it is usually correctly stated that the threshold S/N of the FM demodulator varies directly with the noise bandwidth of the i-f filter that shapes the input noise spectrum. However, it is unjustly stated that the (noise) bandwidth of the i-f filter varies directly with the frequency deviation of the expected signal, and hence the noise threshold of the conventional FM demodulator varies ultimately directly with the deviation ratio of the expected FM signal.

Actually, it can be shown in a number of ways that the bandwidth of a given type of filter varies with the square-root of the frequency deviation of the expected signal, if the tolerance on distortion and the type of filter are held fixed. In fact, direct analysis[1] in terms of series expansion of the response of a linear filter to an FM excitation shows that for a given type of filter and a specified tolerance on acceptable distortion the filter bandwidth (BW) is given by

$$(BW) = K\sqrt{|\dot{\omega}_i(t)|_{max}} \tag{3.1}$$

where $\dot{\omega}_i(t) = d\omega_i(t)/dt$, and $\omega_i(t)$ is the instantaneous frequency of the input FM excitation. The constant K is determined by the filter type and the tolerance on distortion. Thus, if

$$\omega_i(t) = \Delta\Omega\, g(t), \quad |g(t)|_{max} = 1 \tag{3.2}$$

then

$$|\dot{\omega}_i(t)|_{max} = \Delta\Omega\, |\dot{g}(t)|_{max}$$

and

$$(BW) \propto \sqrt{\Delta\Omega} \tag{3.3}$$

The proportionality (3.3) can also be shown as follows. We reason that as the FM signal frequency sweeps across the filter passband, the filter response will depend upon the sweep rate. In particular, if the sweep rate is such that the signal lingers within the nominal passband of the filter for a length of time equal to k time constants of the filter, then if $k > 2$ or 3, the filter response should build up to a near-steady-state pattern, the desired closeness of the pattern to the steady-state form dictating how high k should be. But

$$\text{(filter time constant)} = k_1/(BW), \quad k_1 = \text{const.}$$

and

$$\text{(maximum sweep rate)} = k_2\,\Delta\Omega, \quad k_2 = \text{const.}$$

Consequently, for a specified quality of quasi-static response, we must have

$$(BW)/k_2\,\Delta\Omega = \text{signal dwell time within the nominal passband of the filter}$$

$$= kk_1/BW)$$

whence

$$(BW) \propto \sqrt{\Delta\Omega} \quad , \text{ Q.E.D.}$$

It is a simple numerical exercise to show that use of the commonly employed formula

$$(BW) = 2\,(\Delta\Omega + \omega_m) \tag{3.4}$$

where ω_m = highest modulation frequency, results in a level of modulation distortion that varies with $\Delta\Omega$, for a fixed type of filter.

It has been argued that using the narrower bandwidth indicated by (3.3) would cause the FM signal to partially sweep the falling skirts of the filter response, thus causing its amplitude to dip relative to the noise over parts of the modulation cycle. But the important fact is that the total noise passed by the narrower filter is proportionately lower than the total noise passed by a filter designed in accordance with Eq. (3.4), so that a net reduction in the noise threshold would be perceived in a conventional FM demodulator.

In fact, the point may be stretched farther to suggest that one may consider designing the i-f filter so that its bandwidth is reduced at very low signal levels so that distortion in the filter is traded for a lower demodulator noise threshold. In many applications signal distortion is more palatable than the considerably sharper degradation encountered below the demodulator noise threshold.

The above observations are of fundamental importance in the evaluation and comparison of the threshold performance of various FM demodulators relative to the performance of a conventional FM demodulator.

3.2 The PLL Demodulator

In the study of the ultimate capabilities of PLL demodulators, two questions are of paramount interest in this report. The first is the absolute lower limit on the value of closed-loop bandwidth, the second is the modeling of PLL demodulators for analytic purposes. Both of these questions will be reviewed here for support of later arguments.

3.2.1 Basic Models[2]

The basic functional block diagram of a phase-locked loop (PLL) demodulator is shown in Fig. 1. The signal plus noise at the output of the i-f amplifier may be expressed as

$$e_{if}(t) = [E_s + x_{c,if}(t)] \cos [\omega_{if} t + \psi(t)] - x_{q,if}(t) \sin [\omega_{if} t + \psi(t)] \qquad (3.5)$$

in which

E_s = constant signal amplitude

$\psi(t)$ = instantaneous phase fluctuations of desired signal,
- $\propto$ (message time function) in ϕM,
- $\propto$ (integral of message time function) in FM,

and the noise is represented as the sum of two components that are instantaneously cophasal and in quadrature with the modulated signal and have instantaneous amplitudes $x_{c,if}(t)$ and $x_{q,if}(t)$, respectively. The VFO oscillation may be expressed as

$$e_{osc}(t) = E_{osc} \cos [\omega_{osc} t + \phi_{osc}(t)] \qquad (3.6)$$

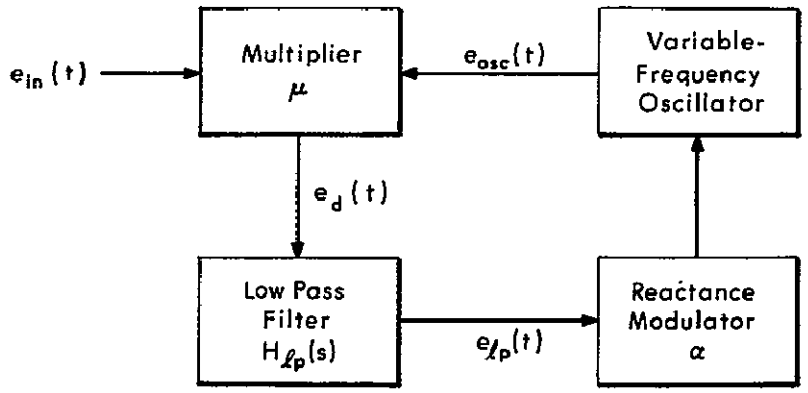


Fig. 1 Basic functional block diagram of phase-locked loop demodulator.

The oscillator instantaneous phase fluctuations, $\phi_{osc}(t)$, in general combine fluctuations due to desired signal modulation, attendant noise and nonlinear interactions of signal and noise. Under the assumption of stable feedback operation, and in the absence of spurious by-products in the output of the multiplier, we have, with $\omega_{if} = \omega_{osc}$,

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc}(t) = \left\{\left[1 + \frac{1}{E_s}x_{c,if}(t)\right]\sin\left[\psi(t) - \phi_{osc}(t)\right] + \frac{1}{E_s}x_{q,if}(t)\cos\left[\psi(t) - \phi_{osc}(t)\right]\right\} \boxtimes h_{\ell p}(t) \qquad (3.7)$$

where $\boxtimes$ denotes convolution and $h_{\ell p}(t)$ is the impulse response of the low-pass filter. The existence of a feedback steady-state condition in which Eq. (3.7) gives an adequate description of loop performance requires that

$$|\Delta\phi| \triangleq |\psi(t) - \phi_{osc}(t)| \leq \pi/2 \qquad (3.8)$$

Equation (3.7), subject to condition (3.8), suggests the model shown in Fig. 2 for PLL operation. If $|\Delta\phi| \ll 1$ almost all of the time, then the sine term may be approximated by its argument and the cosine term by unity. In many practical situations, it is sufficient to require $|\Delta\phi| \leq \sqrt{0.6}$ almost all of the time in order to permit the approximation $\sin\Delta\phi \approx \Delta\phi$, and $|\Delta\phi| \leq \sqrt{0.2}$ almost all of the time, in order to permit the approximation $\cos\Delta\phi \approx 1$. In such situations, the distortion ignored by these approximations is considered to fall below tolerable limits. (This, however, is not true in general, especially in demodulating multichannel FDM/FM signals.)

It is important to note that the condition for approximating $\cos\Delta\phi$ by unity is more stringent than that of approximating $\sin\Delta\phi$ by $\Delta\phi$. Consequently,

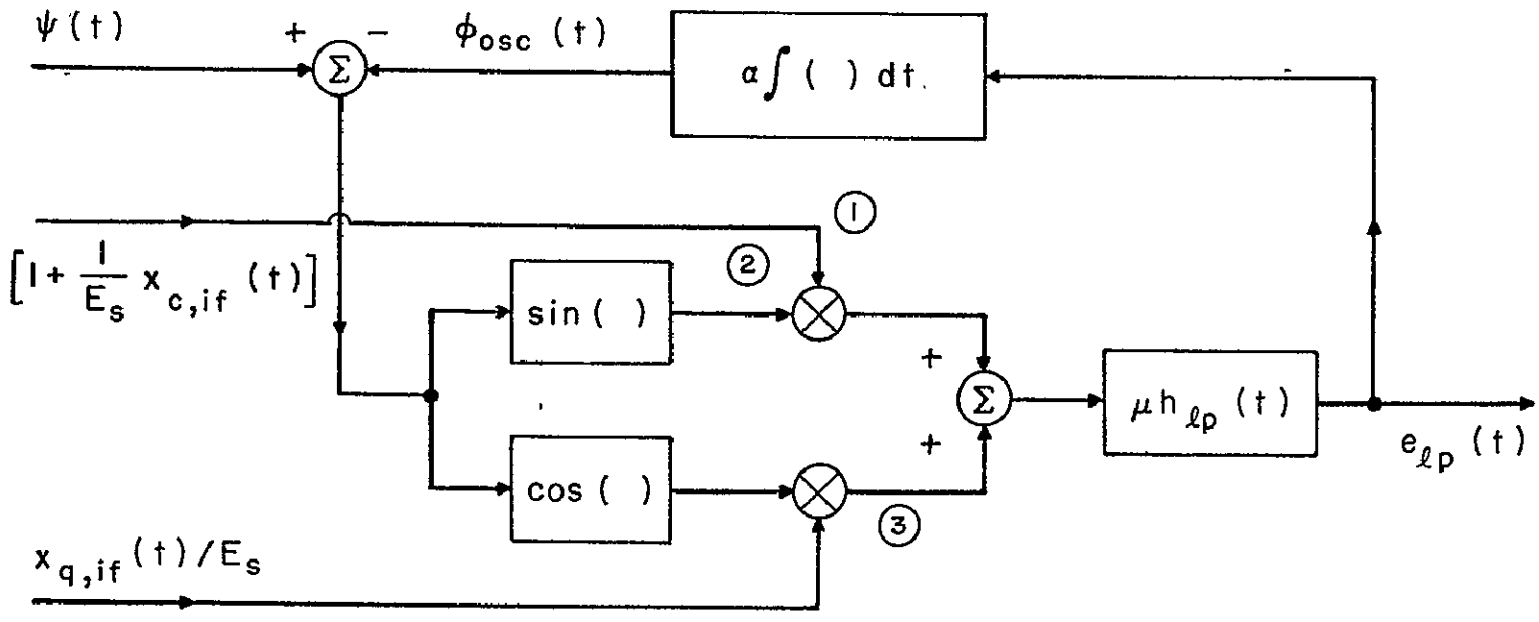


Fig. 2 Model of PLL demodulator in the absence of pre-limiting of the sum of signal and noise, and under the assumptions of stable loop operation, and $|\psi(t) - \phi_{osc}(t)| \leq \pi/2$.

if cos $\Delta\phi$ is approximated by unity, then it follows necessarily that $\sin \Delta\phi \approx \Delta\phi$ and $\Delta\phi$ cannot be allowed to exceed $\sqrt{0.2}$ too frequently.

Thus, approximation of the model of Fig. 2 by the linear model of Fig. 3 requires that the following conditions be satisfied:

(a) The factor at point (1) in Fig. 2 is approximately equal to unity. This means that the cophasal component of the i-f noise must be negligible, which is practically the case if

$$(S/N)_{if} \leq 5 \text{ (or 7 db)} \tag{3.9}$$

(b) $\sin [\psi(t) - \psi_{osc}(t)] \equiv \sin \Delta\phi \approx \Delta\phi$

and

$$\cos \Delta\phi \approx 1$$

These approximations hold if

$|\Delta\phi| \ll 1$ almost all of the time (specifically, $|\Delta\phi| \leq \sqrt{0.2}$) The condition $|\Delta\phi| \ll 1$ reduces the time function at point (2) in Fig. 2 to $\psi(t)-\phi_{osc}(t)$ and that at point (3) to $x_{q,if}(t)/E_s$. (3.10)

When the linear model holds, the output of the low-pass filter is given by the inverse transform of

$$E_{\ell p}(s) = s\Phi_{osc}(s)/\alpha = \frac{\mu H_{\ell p}(s)}{s + \mu\alpha H_{\ell p}(s)} \left[s\Psi(s) + \frac{1}{E_s} sX_{q,if}(s) \right] \tag{3.11}$$

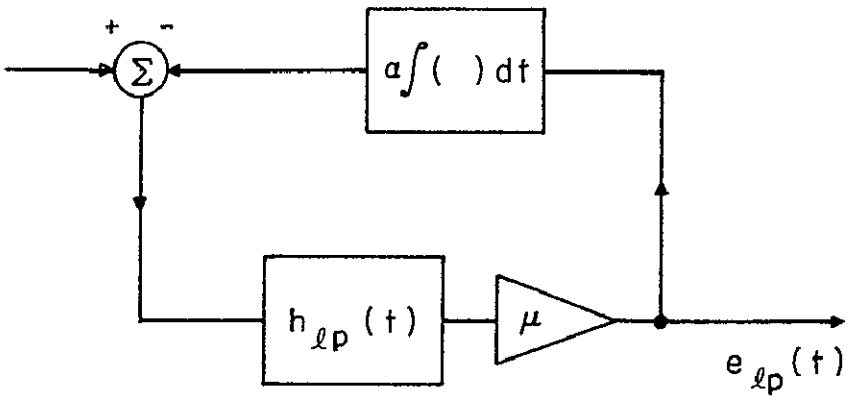


Fig. 3 Baseband linear model of phase-locked loop demodulator operating upon sum of FM signal plus noise. It holds only subject to the conditions

$$|\psi(t) - \phi_{osc,s}(t)| \ll 1, \quad \mu\alpha < \infty$$

and (input S/N in the i-f bandwidth) $\geq$ 7 db.

Thus the use of a low-pass filter that does not cut off abruptly enables frequencies outside the minimum necessary band to be incompletely attenuated, thereby widening not only the equivalent noise bandwidth of the low-pass filter itself, but also the equivalent closed-loop bandwidth for the transmission of instantaneous frequency or phase perturbations from the input signal terminal to the VFO. Low-distortion PLL operation requires that the desired baseband waveform $\dot{\psi}(t)$ be transmitted to the instantaneous frequency of the VFO with negligible distortion. But this does not necessarily mean that the loop low-pass filter need pass the desired modulation, $\dot{\psi}(t)$, with negligible distortion because, within the loop, this filter operates on

$$\frac{\mu s \Psi(s)}{s + \mu\alpha H_{\ell p}(s)}$$

which may in general be a linearly predistorted form of $s\Psi(s)$, but the feedback factor $\mu\alpha$ must be such that $H_{\ell p}(s)$ will provide proper linear compensation so that the resultant system function

$$H_{eq}(s) \equiv \frac{\mu\alpha H_{\ell p}(s)}{s + \mu\alpha H_{\ell p}(s)} \tag{3.12}$$

will yield $\dot{\psi}(t)$ with negligible distortion. Under these conditions, Eq. (3.11) can be simplified to the form

$$s\Phi_{osc}(s) \approx s\Psi(s)e^{-\tau_{d,eq}s} + \frac{1}{E_s} sX_{q,if}(s)\, H_{eq}(s)$$

in which $\tau_{d,eq}$ is the time delay in distortionless transmission through $H_{eq}(s)$. Inverse transformation yields

$$\dot{\phi}_{osc}(t) \approx \dot{\psi}(t - \tau_{d,eq}) + \frac{1}{E_s} \dot{x}_{q,if}(t) \otimes h_{eq}(t) \qquad (3.13)$$

This shows that under conditions of high input S/N ratio, the phase-locked loop simulates the bandpass analog of its equivalent closed-loop instantaneous phase or frequency low-pass transfer function, with its center frequency automatically and continuously adjusted to follow closely the instantaneous frequency of the input signal, provided that the closed-loop equivalent low-pass filter can pass the signal modulation function with negligible distortion and conditions (3.9) and (3.10) are satisfied.

It is important to point out that as long as condition (3.10) is satisfied to within acceptable tolerances, the condition that $H_{eq}(s)$ pass the desired modulation waveform $\psi(t)$ with negligible distortion is not necessary for proper FM demodulation. In fact "optimization" of loop performance with respect to both noise and distortion above the threshold of linear variation of output S/N ratio with i-f S/N ratio may result in non-negligible linear distortion of $\psi(t)$ by $H_{eq}(s)$. Such distortion can, however, be reduced to within tolerable bounds by linear compensation, or "equalization," of the offending non-uniformities in the characteristics of $H_{eq}(s)$ by means of an appropriate low-pass filter after the loop. Such a filter is also generally desirable from the viewpoint of baseband noise filtering because $H_{eq}(s)$ will usually have a larger noise bandwidth than is necessary in proper overall low-pass filtering of the baseband waveform.

If a perfect bandpass limiter operates on the sum of signal and noise before it is applied to the APC loop, Eq. (3.7) is replaced by

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc}(t) = h_{\ell p}(t) \otimes \sin\left[\psi(t) + \theta_{n,if}(t) - \phi_{osc}(t)\right] \tag{3.14}$$

where

$$\theta_{n,if}(t) = \tan^{-1}\frac{x_{q,if}(t)}{E_s + x_{c,if}(t)} \tag{3.15}$$

This suggests the model shown in Fig. 4(a). Approximation of this model by a linear model requires that

$$\left|\psi(t) + \theta_{n,if}(t) - \phi_{osc}(t)\right| \ll 1 \text{ almost all of the time.} \tag{3.16}$$

Moreover, if only the quadrature component of the i-f noise is to be included in the input, then we must also have

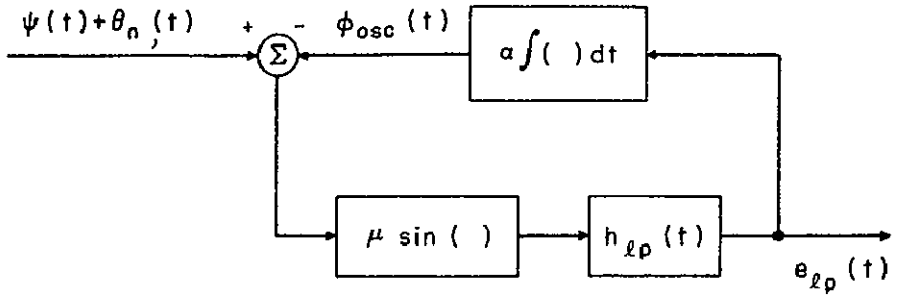
$$\theta_{n,if}(t) \approx \frac{1}{E_s}\dot{x}_{q,if}(t) \tag{3.17}$$

which holds provided that $(S/N)_{if}$ exceeds or equals a "practical" threshold of 10 db.

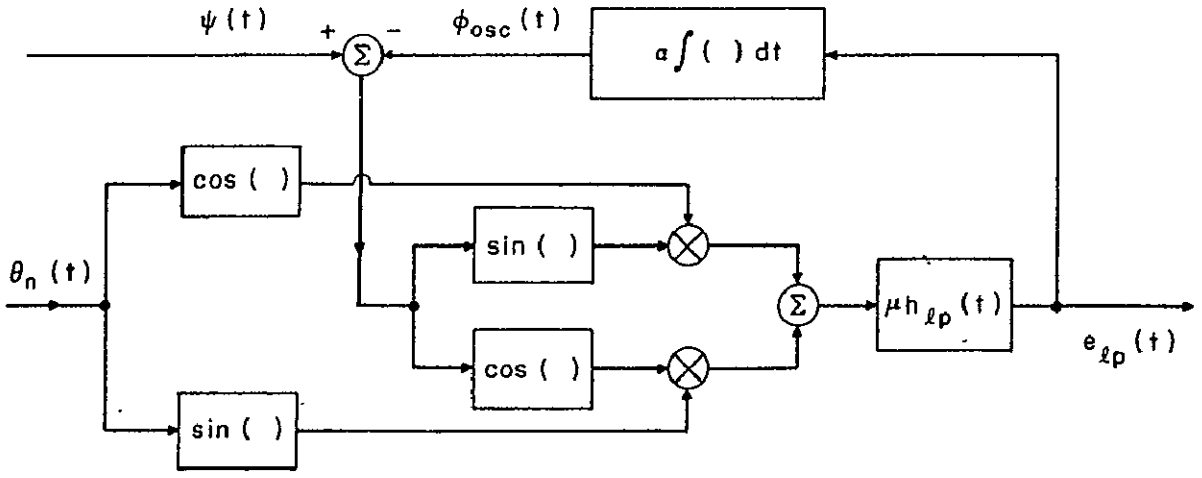
Alternately, Eq. (3.14) may be written as

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc}(t) = h_{\ell p}(t) \otimes \Big\{ \sin\left[\psi(t) - \phi_{osc}(t)\right] \cos\theta_{n,if}(t) + \cos\left[\psi(t) - \phi_{osc}(t)\right] \sin\theta_{n,if}(t) \Big\} \tag{3.18}$$

which suggests the model in Fig. 4(b). Again, linearization requires that condition $|\Delta\phi| \ll 1$ and $(S/N)_{if} \geq 10$ db.



(a)



(b)

Fig. 4 Models of PLL demodulator when the input is the amplitude-limited resultant of signal and noise, under the assumptions of stable loop operation and $|\psi(t) - \phi_{osc}(t)| \leq \pi/2$.

3.2.2 Absolute Lower Limit on Closed-Loop Noise Bandwidth

Let us now consider the situation in which the loop lowpass filter is assumed to have uniform transmission over $B_{\ell p}(<< \mu\alpha)$ rad/sec--the frequency range occupied by the baseband modulation waveform g(t) --with a constant delay τ_d sec and with very large attenuation outside the stated bandwidth. It is well-known that such an attenuation characteristic cannot be realized physically without a correspondingly large delay that is approximately constant only over a fraction of the passband near the center frequency but rises rapidly near the edges of the band. Such a large delay can be shown to be totally unacceptable for realizing a stable feedback system in which the PLL can lock to an FM signal, and yield a negligibly distorted replica of the modulation function in its output. But the idealizations we make concerning constant low delays associated with very sharply selective attenuation characteristics are merely intended to be a crude imitation of physical systems that helps in determining the absolute lower bound on the closed-loop noise bandwidth.

Thus, with the input excitation described by Eq. (3.5) and the oscillation expressed as in Eq. (3.6), then, in the absence of spurious by-products in the output of the multiplier (or "phase detector"), the low-frequency components applied to the lowpass filter are

$$\mu\left[1 + \frac{x_{c,if}(t)}{E_s}\right]\sin\Delta\phi + \frac{1}{E_s}\,\mu\, x_{q,if}(t)\cos\Delta\phi \qquad (3.19)$$

A feedback steady-state condition exists, in which (3.19) properly describes the low-frequency components applied to the lowpass filter, provided that $|\Delta\phi| \leq \pi/2$. If we assume that $\Delta\phi < \sqrt{0.2}$, then (3.9) can be simplified to the form

$$\mu\left[1 + \frac{1}{E_s} x_{c,if}(t)\right] \Delta\phi + \frac{\mu}{E_s} x_{q,if}(t) \tag{3.20}$$

The output of the lowpass filter will therefore be

$$e_{\ell p}(t) \approx \mu\left[\psi(t - \tau_{\ell p}) - \phi_{osc}(t - \tau_{\ell p})\right] + \frac{\mu}{E_s}\left\{x_{c,if}(t)\left[\psi(t) - \phi_{osc}(t)\right]\right\} \otimes h_{\ell p}(t) + \frac{\mu}{E_s} x_{q,if}(t) \otimes h_{\ell p}(t) \tag{3.21}$$

If the baseband bandwidth is a small fraction of the i-f bandwidth, the second term in Eq. (3.21) can be approximated by

$$\frac{\mu}{E_s}\left[\psi(t) - \phi_{osc}(t)\right]\left[x_{c,if}(t) \otimes h_{\ell p}(t)\right] \tag{3.22}$$

When expressed in this form, it becomes clear that the second term in Eq. (3.21) is practically negligible if the ratio of mean squared signal to mean squared noise contained within twice the idealized lowpass filter bandwidth, centered about the instantaneous frequency of the input signal, exceeds or equals 5 ($\approx$7 db). Expression (3.20) shows that, in general, the second term in Eq. (3.21) can for practical purposes be neglected as long as the i-f S/N ratio exceeds or equals 7 db.

If the second term in Eq. (3.21) can be neglected, we have

$$e_{\ell p}(t) \approx \mu \left[\psi(t - \tau_{\ell p}) - \phi_{osc}(t - \tau_{\ell p}) \right] + \frac{\mu}{E_s} x_{q,if}(t) \boxtimes h_{\ell p}(t) \qquad (3.23)$$

Thus, under high S/N conditions we have

$$\dot{\phi}_{osc}(t) = \alpha e_{\ell p}(t)$$

$$\approx \mu\alpha\psi(t - \tau_{\ell p}) - \mu\alpha\phi_{osc}(t - \tau_{\ell p}) + \frac{\mu\alpha}{E_s} x_{q,if}(t) \boxtimes h_{\ell p}(t) \qquad (3.24)$$

whence

$$s\Phi(s) \approx \Psi(s)\, e^{-s/\mu\alpha} + \frac{1}{E_s}\, s X_{q,if}(s)\, e^{-s/\mu\alpha} \cdot H_{\ell p}(s)\, e^{\tau_{\ell p} s}$$

or

$$\dot{\phi}_{osc}(t) \approx \Delta\Omega g(t - 1/\mu\alpha) + \frac{1}{E_s} x_{q,if}(t - 1/\mu\alpha) \boxtimes h_{\ell p}(t + \tau_{\ell p}) \qquad (3.25)$$

This shows that the delay τ_d required for yielding FM demodulation by the phase-locked loop with a sharp cutoff lowpass filter is given by $1/\mu\alpha$, the time constant of the loop excluding the normalized, idealized lowpass filter. In order to guarantee that this delay is small but nonzero so that proper FM demodulation can result, the loop gain $\mu\alpha$ should be large (but not infinite).

The result expressed in Eq. (3.25) shows that the phase-locked loop with a rectangular lowpass filter within the loop has a noise bandwidth equal to the bandwidth of the lowpass filter. But the minimum permissible value of this bandwidth is twice $B_{\ell p}$ (in order to include positive and negative frequencies) where $B_{\ell p}$ is the (positive-frequency) bandwidth

of the desired message. Consequently, we may conclude that the absolute lower bound on closed-loop noise bandwidth is $2B_{\ell p}$, and, hence, the absolute upper bound on the threshold reduction factor achievable with a phase-locked loop is

$$2\cdot(B_{if}/2B_{\ell p}) = B_{if}/B_{\ell p} \tag{3.26}$$

The factor of 2 in (3.26) takes into account the fact the threshold of linear variation of baseband S/N ratio with i-f S/N ratio for a PLL demodulator is reached when the S/N ratio in the oscillation noise bandwidth equals 5 ($\approx$7 db).

3.3 The FCF Demodulator

The functional diagram of an FCF demodulator is shown in Fig. 5. We assume that the closed-loop system can lock properly to the desired signal in the absence of noise, and let $e_{if}(t)$ and $e_{osc}(t)$ be described by (3.5) and (3.6), respectively. The FCF loop is assumed to be driven with no prior amplitude limiting of signal plus noise. Under these conditions, we further assume that the mixer delivers in the nominal passband of the loop bandpass filter only the beat components

$$[E_s + x_{c,if}(t)] \cos [\omega_{bp} t + \psi(t) - \phi_{osc}(t)]$$

$$- x_{q,if}(t) \sin [\omega_{bp} t + \psi(t) - \phi_{osc}(t)]$$

in which $\omega_{bp} \equiv \omega_{if} - \omega_{osc}$ = nominal center frequency of the bandpass filter within the loop. If an amplitude-insensitive linear FM demodulator and a

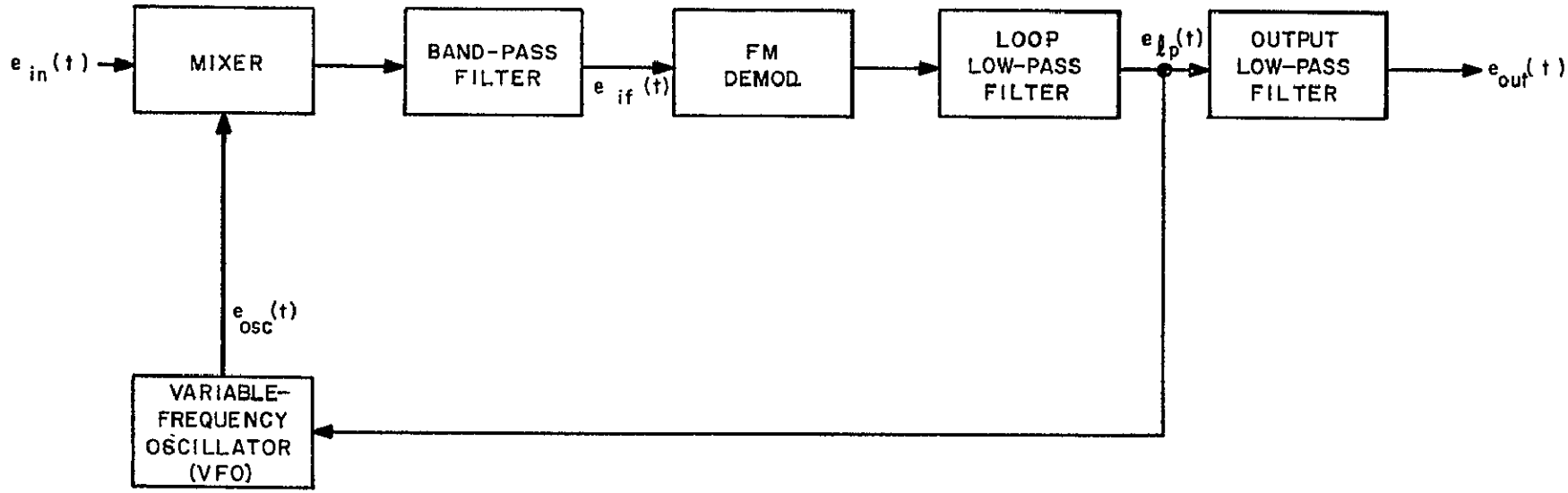


Fig. 5 Basic functional block diagram of frequency-compressive feedback demodulator.

linear feedback from the discriminator output to the instantaneous frequency of the VFO are assumed, we can show that

$$\phi_{osc}(t) = k_d \beta_{fb} \left[\tan^{-1} \frac{A_q(t) \otimes h_{LP}(t)}{A_c(t) \otimes h_{LP}(t)} \right] \otimes h_{\ell p}(t) \qquad (3.27)$$

where

$$A_c(t) \equiv \left[1 + \frac{1}{E_s} x_{c,if}(t) \right] \cos \left[\psi(t) - \phi_{osc}(t) \right] - \frac{1}{E_s} x_{q,if}(t) \sin \left[\psi(t) - \phi_{osc}(t) \right]$$

$$A_q(t) \equiv \left[1 + \frac{1}{E_s} x_{c,if}(t) \right] \sin \left[\psi(t) - \phi_{osc}(t) \right] + \frac{1}{E_s} x_{q,if}(t) \cos \left[\psi(t) - \phi_{osc}(t) \right]$$

and $h_{LP}(t)$ is the impulse response of the low-pass analog of the bandpass filter.

3.3.1 Basic Models[2]

Expression (3.27) for $\phi_{osc}(t)$ can be used to derive lowpass equivalent models of the FCF operation upon the input signal plus noise for arbitrary input SNR's. We illustrate this with a number of important examples.

Consider first the situation in which the noise is so weak that it can be ignored entirely (i.e., the noiseless case). Here (3.27) is simplified to

$$\phi_{osc}(t) = k_d \beta_{fb} \, h_{\ell p}(t) \boxtimes \tan^{-1} \frac{h_{LP}(t) \boxtimes \sin[\psi(t) - \phi_{osc}(t)]}{h_{LP}(t) \boxtimes \cos[\psi(t) - \phi_{osc}(t)]} \qquad (3.28)$$

This suggests a general nonlinear baseband model of the FCF operation upon signal modulation. It is clear, however, that linear-approximation models of the FCF operation upon the signal modulation can be justified under either of the two conditions:

(a) $|\psi(t) - \phi_{osc}(t)| \ll 1$ (or ≤ 0.2 for many practical purposes) which allows the approximation

$$\tan^{-1} \frac{h_{LP}(t) \boxtimes \sin[\psi(t) - \phi_{osc}(t)]}{h_{LP}(t) \boxtimes \cos[\psi(t) - \phi_{osc}(t)]}$$

$$\approx h_{LP}(t) \boxtimes [\psi(t) - \phi_{osc}(t)] \qquad (3.29)$$

or

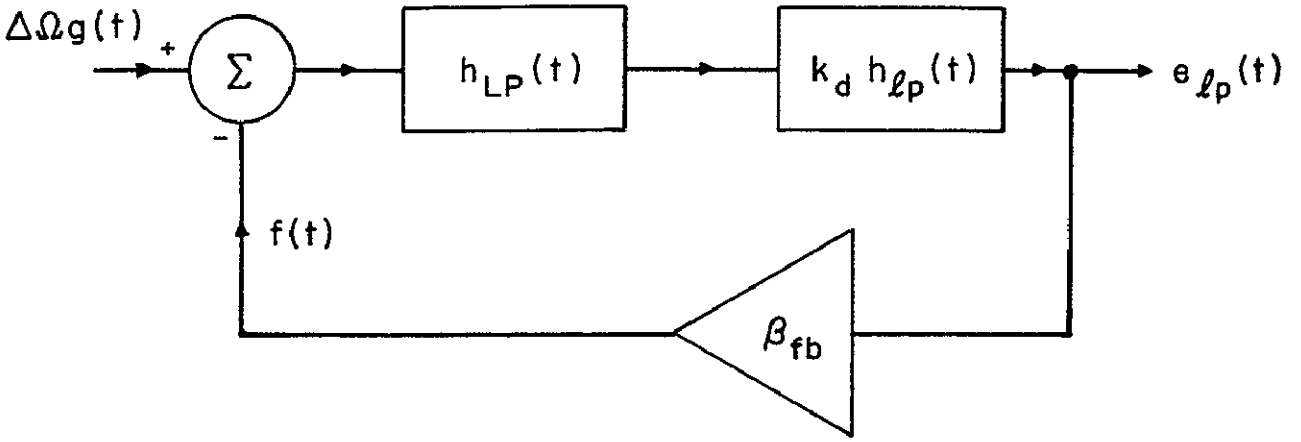
(b) the response of the loop bandpass filter to the frequency-compressed signal is quasi-stationary with negligible nonlinear distortion of the modulation, which allows the approximation

$$\tan^{-1} \frac{h_{LP}(t) \boxtimes \sin[\psi(t) - \phi_{osc}(t)]}{h_{LP}(t) \boxtimes \cos[\psi(t) - \phi_{osc}(t)]}$$

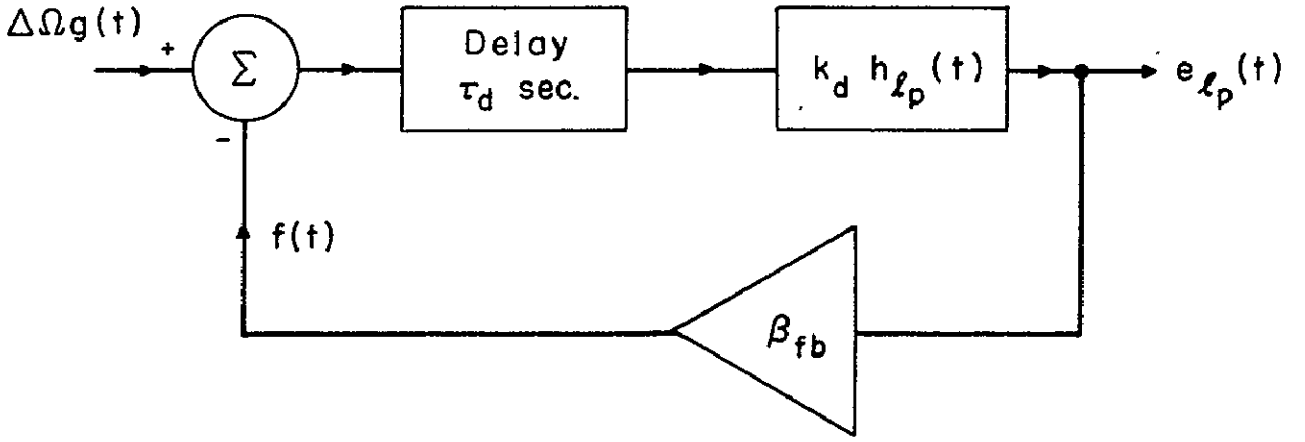
$$\approx \tan^{-1} \frac{\sin[\psi(t - \tau_d) - \phi_{osc}(t - \tau_d)]}{\cos[\psi(t - \tau_d) - \phi_{osc}(t - \tau_d)]}$$

$$\approx \psi(t - \tau_d) - \phi_{osc}(t - \tau_d) \qquad (3.30)$$

Condition (a) leads to the model of Fig. 6(a); condition (b) leads to 6(b). The two models differ only in the equivalent lowpass analog of the loop bandpass filter



(a) Subject to condition that the FCF reduces the phase deviation at the input of the loop bandpass filter to a value much smaller than unity (0.2 rad or less).



(b) For operation on signal modulation. Valid only when the loop bandpass filter responds to the frequency-compressed signal in a quasi-stationary manner, with negligible distortion, and delivers replica of compressed modulation function delayed τ_d sec.

Fig. 6 Linear models of FCF demodulator.

With reference to Fig. 6, the closed-loop system function is given by

$$H_{eq}(s) = \frac{k_b \beta_{fb} H_{LP}(s) H_{\ell p}(s)}{1 + k_d \beta_{fb} H_{LP}(s) H_{\ell p}(s)} \tag{3.31}$$

for the model of Fig. 6(a), and

$$H_{eq}(s) = \frac{k_d \beta_{fb} e^{-\tau_d s} H_{\ell p}(s)}{1 + k_d \beta_{fb} e^{-\tau_d s} H_{\ell p}(s)} \tag{3.32}$$

for the model of Fig. 6(b).

The frequency modulation of the signal at the input of the loop bandpass filter is given by the inverse transform of

$$H_{eq,c}(s)\, s\Psi(s) \tag{3.33}$$

in which

$$H_{eq,c}(s) \triangleq 1 - H_{eq}(s) \tag{3.34}$$

is the complement of the closed-loop system function $H_{eq}(s)$. In general, $H_{eq,c}(s)$ may cause linear distortion of the desired modulation, but such distortion must not exceed certain low bounds if proper operation is to be maintained in the presence of noise down to and below the threshold of a conventional FM demodulator.

When the linear models cannot be justified, the nonlinear effects of the FCF operation introduced in the response of the loop bandpass filter can be computed with the aid of Eq. (3.27).

Linear and nonlinear approximation models of the FCF operation can also be developed when the noise cannot be ignored. Of particular interest are the models that apply under conditions of high input S/N ratio. We distinguish the two situations:

$$|\psi(t) + \theta_{n,if}(t) - \phi_{osc}(t)| < 1 \tag{3.35}$$

almost all of the time and

$$|\psi(t) + \theta_{n,if}(t) - \phi_{osc}(t)| \geq 1 \tag{3.36}$$

a considerable fraction of the time.

In the first of these situations,

$$\frac{1}{k_d \beta_{fb}} \dot{\phi}_{osc}(t) \approx [\psi(t) + \theta_{n,if}(t) - \phi_{osc}(t)] \otimes h_{BP}(t) \otimes h_{\ell p}(t)$$

which establishes the applicability of the linear model of Fig. 6(a) with $\psi(t) + \theta_{n,if}(t)$ as representative of input signal plus noise. This linear model yields a practically satisfactory approximation to the actual S/N mean-square ratio performance provided that

$$\theta_{n,if}(t) \approx \frac{1}{E_s} x_{q,if}(t) \tag{3.37}$$

to within a first-order mean-square contribution from $(1/E_s)\, x_{c,if}(t)$ that is at least 10 db below that of $(1/E_s)\, x_{q,if}(t)$ — i.e., as long as the S/N ratio in the pre-FCF bandwidth exceeds or equals about 12 db.

For the situation in which (3.36) applies, we recall that under conditions of high input S/N ratio, the effects of signal modulation and of noise upon the output can be computed separately. Thus, if we first ignore the noise, then with

$$|\psi(t) - \phi_{osc}(t)| > 0.2$$

the model of Fig. 6(b) applies subject to the condition that the loop bandpass filter response reproduce the compressed modulation with negligible distortion. Next, ignoring the signal modulation, we observe that with

$$|\theta_{n,if}(t)| \approx \left|\frac{1}{E_s} x_{q,if}(t)\right|$$

well below unity most of the time (i.e., with the pre-FCF S/N ratio exceeding about 12 db), the linear model of Fig. 6(a) will approximate closely the effect of FCF upon the FM noise. Thus, we have one model for signal modulation (Fig. 6(b)) and another for noise (Fig. 6(a)).

The preceding results show that the FCF demodulator operating upon the sum of signal plus noise simulates a linear tracking filter given by the bandpass analog of the equivalent closed-loop lowpass transfer function $H_{eq}(s)$ defined in Eq. (3.31), with its center frequency continuously and automatically adjusted to track the instantaneous frequency of the input signal, provided that the desired signal modulation is reproduced with negligible distortion by the closed loop system, and $(S/N)_{if} \geq$ 12 db.

3.3.2 Absolute Lower Limit on Closed-Loop Noise Bandwidth

Let us now consider the situation in which the loop bandpass and lowpass filters in Fig. 5 are assumed to have uniform transmission over B_{BP} and $B_{\ell p}$ rad/sec, respectively, with constant delays of τ_d and τ_ℓ sec, and with very large attenuation outside the stated bandwidths. Assume that relatively weak noise is superimposed on an input unmodulated carrier.

Now, since

a) the noise in the passband of the loop bandpass filter is much weaker than the signal;

b) the open-loop lowpass filter action, by assumption, restricts the transmission to a bandwidth that is much narrower than the bandwidth of the receiver stages that drive the FCF demodulator; and

c) the bandwidth of the loop bandpass filter is sufficient to enable this filter to respond quasi-stationarily to the weak frequency fluctuations by the noise (i.e., $H_{BP}(j\omega)$ is essentially uniform — with unit amplitude and phase slope τ_d, say — over the range within which the noise-deviated instantaneous frequency of the local oscillation is bounded almost all of the time);

we have at the output of the loop bandpass filter

$$e_{BP}(t)/K_{if}E_{osc} \approx \left[E_s + x_{c,rf}(t) \otimes h_{LP}(t)\right] \cdot \cos\left[\omega_{if}t - \phi_{osc}(t-\tau_d)\right] - \left[x_{q,rf}(t) \otimes h_{LP}(t)\right] \cdot \sin\left[\omega_{if}t - \phi_{osc}(t-\tau_d)\right] \tag{3.38}$$

In response to $e_{BP}(t)$, the amplitude-insensitive FM demodulator delivers to the loop lowpass filter an excitation proportional to

$$\frac{1}{E_s}\dot{x}_{q,rf}(t) \otimes h_{LP}(t) - \phi_{osc}(t-\tau_d) \tag{3.39}$$

If the loop lowpass filter has an impulse response denoted by $h_{\ell p}(t)$, then

$$e_{\ell p}(t) = \frac{k_d}{E_s} \dot{x}_{q,rf}(t) \otimes h_{LP}(t) \otimes h_{\ell p}(t) - k_d \dot{\phi}_{osc}(t - \tau_d - \tau_{\ell p})$$

$$= \frac{k_d}{E_s} \dot{n}_{\ell p}(t) - k_d \dot{\phi}_{osc}(t - \tau_d - \tau_{\ell p}) \qquad (3.40)$$

wherein we have defined

$$n_{\ell p}(t) = x_{q,rf}(t) \otimes h_{LP}(t) \otimes h_{\ell p}(t) \qquad (3.41)$$

We have also chosen to ignore the filtering, except for delay $\tau_{\ell p}$, of $\dot{\phi}_{osc}(t)$ by $h_{\ell p}(t)$ in (3.40) on the basis that $\dot{\phi}_{osc}(t)$ is already a lowpass function that has received sufficient filter action by $h_{\ell p}(t)$ to allow it to pass again through this filter with no modification.

If we set $\phi_{osc}(t) = \beta_{fb} e_{\ell p}(t)$ in (3.40), we obtain

$$E_{\ell p}(s) = \frac{k_d/E_s}{1 + k_d \beta_{fb} e^{-(\tau_d + \tau_{\ell p})s}} \; s N_{\ell p}(s)$$

which, for negligible open-loop delay and/or for (the tacitly assumed) large feedback, leads to

$$e_{\ell p}(t) = \frac{k_d/E_s}{1 + k_d \beta_{fb}} \; n_{\ell p}(t + \tau') \qquad (3.42)$$

$$\tau' \triangleq \frac{k_d \beta_{fb}}{1 + k_d \beta_{fb}} (\tau_d + \tau_{\ell p})$$

Finally, the noise transmitted to the oscillator phase is

$$\phi_{osc}(t) = \frac{k_d/E_s}{1 + k_d \beta_{fb}} \; n_{\ell p}(t + \tau) \tag{3.43}$$

From Eqs. (3.41) and (3.43) it is clear that the closed-loop noise bandwidth of an FCF demodulator with a rectangular lowpass filter within the loop is given by the bandwidth (including positive and negative frequencies) of the lowpass filter (i.e., $2B_{\ell p}$). This is then the absolute lower limit on FCF closed-loop noise bandwidth.

The determination of the noise threshold requires an examination of which of two conditions will break down first:

(i) Operation of the loop discriminator above its threshold; or

(ii) $\overline{\phi^2_{osc,\,n}(t)} \leq 1/20$ (which guarantees that the contribution from the interaction of oscillator phase noise, $\phi_{osc,\,n}(t)$, and input noise is negligible).

If the threshold defined by condition (i) is denoted $(S/N)_1$ and that defined by (ii) is $(S/N)_2$, then

$$(S/N)_1/(S/N)_2 = \frac{2B_{\ell p}}{B_{BP}} < 1 \tag{3.44}$$

which shows that condition (i) breaks down first and, hence, defines the closed-loop threshold. Consequently, the absolute upper bound on the threshold reduction factor achievable with FCF is

$$B_{if} \; / \; B_{BP} \tag{3.45}$$

In comparison with the upper bound for PLL, the upper bound given for FCF by (3.45) is lower by a factor of

$$B_{BP}/B_{\ell p} > 2 \qquad (3.46)$$

The limit of threshold reduction is achieved with FCF using idealized rectangular filters when the frequency compression reduces the necessary bandpass filter bandwidth to a value close to twice the baseband bandwidth. The upper bound on the permissible value of $K_d\beta_{fb}$ and the gradual cutoff encountered (and required) in practical filters restrict the degree to which this limit can be approached.

3.4 Feedforward Band-Dividing Demodulator

One promising band-dividing scheme is proposed in Fig. 7. The instantaneous frequency of the input signal is estimated by a bank of filters or its time-domain equivalent. An FM signal at the estimated instantaneous frequency is then generated. The first mixer takes the frequency difference between the input signal and the estimated signal. If the estimate is reasonably accurate (this point will be expanded later) then the frequency difference will have much smaller excursions than the input frequency. The difference signal may therefore be passed through a relatively narrowband filter, thereby eliminating much of the input noise power. The filtered difference signal is mixed with the estimate a second time, restoring the original excursions in input instantaneous frequency. The result of these operations is, under ideal conditions, a replica of the input FM signal accompanied by a relatively narrow band of noise surrounding the instantaneous frequency of the input signal. A

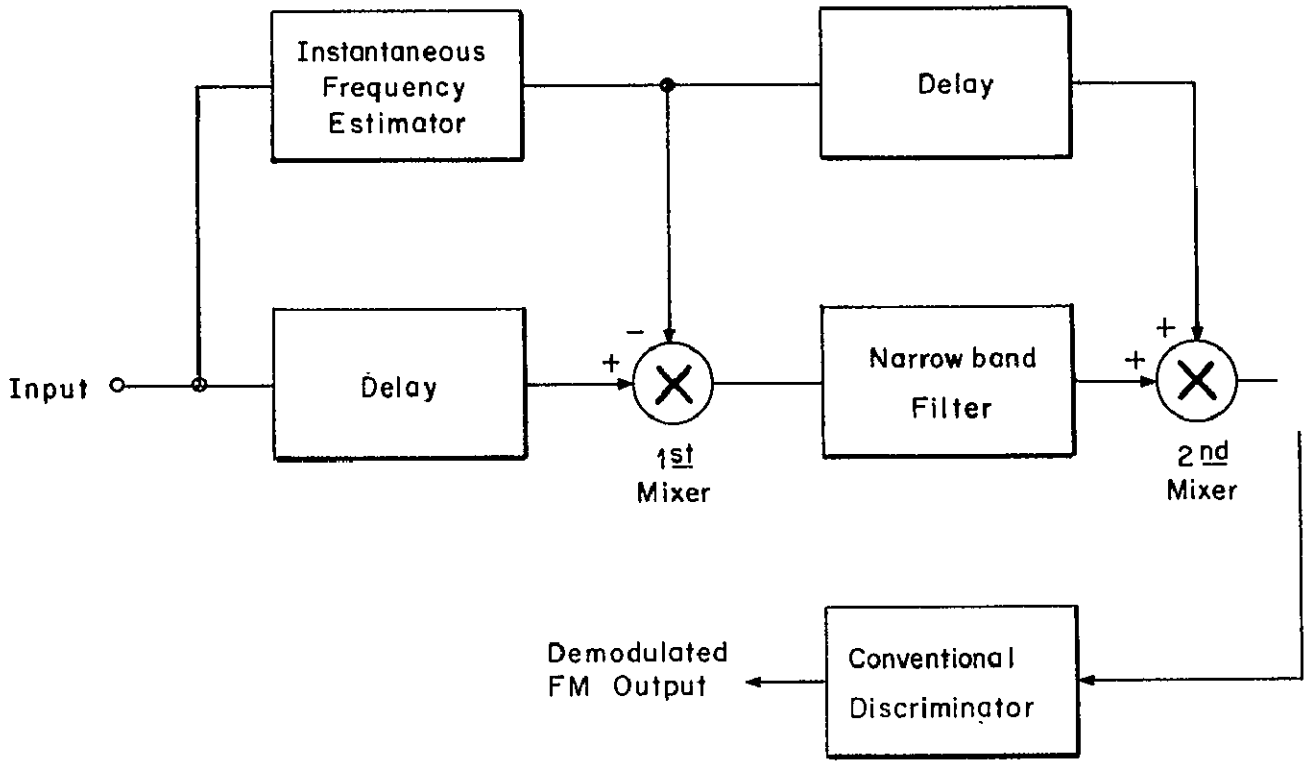


Fig. 7 Band-dividing scheme.

conventional FM demodulator driven by this signal will have a lowered threshold because of the reduced noise power at its input.

The amount of threshold reduction afforded by such a scheme depends upon two major factors:

(1) The ability of the estimator to furnish an error-free estimate of the input frequency within a specific accuracy in the presence of noise, and

(2) The presence of a coherent difference frequency signal which can be narrowband filtered without introducing serious fluctuations in amplitude.

Each of these requirements must be carefully examined. In what follows, we take up the problem of estimation for the various possible forms in which the desired baseband signal can be presented to the FM demodulator at the transmitting end. The coherence requirement is now under study and will be covered in a later report.

3.4.1 Estimator Requirements for Various Message Representations

The band-dividing concept involves a determination of the input instantaneous frequency by a set of contiguous bandpass filters or some other equivalent structure.[3,4,5] There are at least three ways of processing the message prior to frequency modulation:

(1) No processing other than pre-emphasis by a linear filter,

(2) Sampling and quantization into a set of fixed levels, and

(3) Sampling and holding with no quantization.

The required number of filters and the filter bandwidths are different for the three cases.

Consider first the case (No. 1) of an analog presentation of the baseband signal to the FM modulator. At the receiver, the "signal estimator" is made up of a suitably designed band contiguous bandpass filters each followed by an appropriate response measuring system. The important point here is that the (analog) instantaneous frequency must stay within the bandwidth of a given demodulator filter long enough for the filter output to build up to a response closely approximating the steady state, otherwise noise and distortion will dominate. If the filter bandwidth is given by $B_{a,f}$ then the instantaneous frequency must remain within $B_{a,f}$ for no less than $2\pi/B_{a,f}$ sec. Assuming a worst-case situation of sinusoidal FM of index δ at the maximum message frequency, ω_m, the instantaneous frequency is given by

$$\omega_i(t) = \omega_c + \delta\omega_m \sin\omega_m t. \tag{3.47}$$

The maximum instantaneous frequency rate is,

$$\left.\frac{d\omega_i}{dt}\right|_{max} = \delta\omega_m^2 \text{ radians/sec}^2. \tag{3.48}$$

Thus, the value that must be exceeded by the filter bandwidth is determined from

$$\left.\frac{d\omega_i}{dt}\right|_{max} \Delta t = \delta\omega_m^2 \cdot \left(\frac{2\pi}{B_{a,f}}\right) \leq B_{a,f} \tag{3.49}$$

whence

$$B_{a,f} \geq \sqrt{2\pi}\ \omega_m \sqrt{\delta}\ . \qquad (3.50)$$

This simple result shows that the filter bandwidth in general must be considerably greater than the Nyquist rate of $2\omega_m$ and that the excess bandwidth increases with the modulation index. A similar relationship holds for the optimized second order phase-locked loop, and, as was shown in Section 3.1, for the usual i-f amplifier followed by a conventional FM discriminator. On this basis, the band-dividing approach clearly does not offer a worthwhile approach to FM demodulation.

Consider next the case (No. 2) of sampling and quantizing prior to frequency modulation. At the receiver each level of quantization has a corresponding filter. The allowable minimum sampling frequency, f_s, is twice the message bandwidth. Thus, if the highest significant frequency in the message spectrum is slightly less than f_m cps, then the widest allowable time spacing between the periodically timed samples is $1/f_{s,\min} = 1/2\, f_m$ sec. Consequently, the radiated FM signal is made up of a sequence of rectangular pulses $1/2\, f_m$ sec wide, each of which envelopes an r-f signal having one of a finite set of frequencies. It is well-known that such a frequency-hopping signal in the presence of white, gaussian noise is best estimated by means of a bank of filters each matched to one of the rectangular r-f pulses. The normalized magnitude squared of the transmittance of each of these filters (shifted down to $\omega = 0$) is given by

$$\left[\frac{\sin(\omega/2f_s)}{\omega/2f_s} \right]^2 \tag{3.51}$$

The (noise) bandwidth of such a filter is given by the integral of (3.51) between $-\infty$ and $+\infty$, which yields

$$B_{m,f} = 2\pi f_s \text{ rad/sec.} \tag{3.52}$$

For the allowable minimum value of f_s,

$$f_{s,min} = 2f_m$$

the bandwidth $B_{m,f}$ of each matched filter is

$$B_{m,f,min} = 2\pi \cdot 2f_m \tag{3.53}$$

Since number of levels necessary to obtain a low level of quantization noise is usually large, a large number of filters spaced at least by twice the message bandwidth are required. This system therefore must have a deviation equal to the number of quantization levels times the message bandwidth, or a modulation index of N = number of levels. Usually the required index will be impractically large, which makes case No. 2 not particularly interesting.

The third case (No. 3) is demodulation of a signal in which the message has been subjected to a sample-and-hold operation prior to frequency modulation. Here, the instantaneous frequency is constant between samples and the filter bandwidth can be equal to the sampling frequency. The sample-and-hold operation does not destroy the analog message provided that the sampling rate is greater than twice the message bandwidth.

With no quantization employed, the constant values of frequency between pulses can assume any of the values in a continuum that covers the range of expected frequency excursions. Thus, the advantage of pre-fixing a bank of filters matched to a discrete set of pulses — made possible by quantization — vanishes. Only sub-optimum solutions are now practicable. One such solution is to use a bank of filters that are matched to a discrete set of the frequencies in the desired range. The specific selection of frequencies may be made either on an ad hoc basis to be uniformly spaced within the desired range, or to be distributed on the basis of the probability density function of the baseband message process thus favoring the more probable frequency positions. One may also consider a discrete set of adaptive matched filters each of which acts like a phase-locked loop that can be frequency pulled only within a designated frequency sub-interval.

The performance of the estimator in the present case (No. 3) can be analyzed on the basis of determining the probability of error in identifying the frequency position of the signal from observation and comparison of the filter outputs. Such an analysis should take proper account of the fact that the filter most nearly matched to a particular signal pulse will be detuned relative to the actual pulse frequency. The amount of detuning will be distributed in accordance with the conditional probability distribution of the message sample values in the designated range. The effects of "matched" filter mismatch and detuning upon the probability of error are now being studied and will be covered in a future report.

Akima[4] has considered only the demodulation of a signal frequency modulated by a sample-and-hold representation of the baseband message. His results suffer from two significant shortcomings of the analysis:

(a). He erroneously applied error probability formulas, derived by Reiger[6] for the proper matched filter case, to the present case in which the filters can be detuned from the actual frequencies by up to one full baseband bandwidth.

(b) He erroneously assumed that under high SNR conditions the filter bank will estimate the instantaneous frequency within $\pm \omega_m$ where ω_m is the message band limit. This assumption ignores the potentially significant FM transient effects.[1]

Something must further be done to refine the frequency estimates after the estimator; otherwise, there will be a quantization noise which is unacceptably large for FM signals of reasonable modulation indices. Akima proposes to measure the instantaneous frequency at the output of the band-dividing filter having the largest amplitude. His handling of this process and assumptions about its results are inadequate. While this measurement, in principle, could be accomplished, the amount of circuitry involved would be rather large. A more workable alternative is the system proposed here in Fig. 7.

Near the threshold, the estimator in Fig. 7 will occasionally make a mistake in selecting the proper filter because the noise amplitude in some filter will exceed the amplitude of the filter which contains the signal. Under moderately high SNR conditions, the probability of a mis-selection is

given by (N-1) times the probability of error for the corresponding binary symmetric system (2 filters).[6] These mis-selection errors constitute the threshold noise (or "error noise" as defined by Battail) of the band dividing demodulator. While the error rate will depend upon the sampling rate and to a certain extent upon the number of channels, for a given sampling rate and about 10 channels the threshold noise will become significant when the SNR in a bandwidth of $2\omega_m$ drops to about 12 db. (Akima's intrinsic SNR is measured in a bandwidth of ω_m and is therefore 3 db higher).

Therefore, regarding the selection of a representation scheme for the input analog message it is evident that the sampling scheme which permits the smallest possible bandwidth for the estimating filters will give the lowest threshold. Also the sampling scheme must not introduce an excessive amount of quantization noise. Of the sampling schemes considered, the sample-and-hold operation is best in this regard since it gives a minimum estimator bandwidth of $2\omega_m$ regardless of the modulation index. Straightforward analog modulation requires an estimator bandwidth larger than this with the bandwidth increasing as the square root of the modulation index.

3.5 Effects of Non-matched Filtering and the Interpulse Frequency Measurement in Band-Dividing Estimators

Assume a sample-and-hold message frequency modulated on a carrier. The estimator contains a bank of filters, each filter having an impulse response described by,

$$h_n(t) = 2g(t) \cos \omega_n t \tag{3.54}$$

where g(t) is real and lowpass. The filters are not necessarily matched to the input signal pulses. During any one signal pulse, the filter is hit with an input consisting of

$$\begin{aligned} r(t) &= s(t) + n(t) \quad ; \quad t > 0 \\ &= n(t) \quad ; \quad t < 0 \end{aligned} \tag{3.55}$$

where s(t) is the pulse of signal and n(t) is white gaussian noise. The output of the filter at the end of the square signal pulse is

$$e_o(T) = \int_o^T s(t)\, h_n(T-t)\, dt + \int_{-\infty}^T n(t)\, h_n(T-t)\, dt. \tag{3.56}$$

Adopting complex notation the output is

$$\overline{e_o}(t) = \left[\int_o^T g(T-t)\, e^{j(\omega_s - \omega_n)t}\, dt + \overline{V_n} \right] e^{j\omega_n t} \tag{3.57}$$

where $\overline{V_n}$ has the magnitude and phase of the (narrowband) filtered noise in (3.56). The envelope of the filter current is the magnitude of (3.57). The magnitude

of the signal term in (3.57) is essentially the magnitude of the Fourier Transform of the lowpass impulse response from 0 to T evaluated at $\omega_s - \omega_n$ or

$$|\overline{S_{out}}(T)| = \left| \int_0^T g(t)\, e^{-j\omega t}\, dt \right|_{\omega = \omega_s - \omega_n} \qquad (3.58)$$

There will be a degradation in (3.58) over the matched filter case if g(t) is not the impulse response of the matched filter and because of the detuning effect. The probability of error in selecting which filter has the signal will therefore increase. If the inter-channel interference problem is ignored, the degraded S/N ratio simply enters into the usual expressions for the system probability of error since the decision problem remains unchanged. The detuning is, of course, a function of the message and an exact statistical description of the detuning is difficult to obtain. Related to the detuning is the inter-channel interference caused by the response of more than one filter to the signal pulse. This effect will increase the probability that near the threshold a channel adjacent to the signal one is erroneously selected. With a signal term present in more than one filter the decision problem itself is modified and available results no longer apply.

Results for Single-Pole Filters

The matched filter is optimum in the sense that it maximizes the ratio of peak output signal power to average output noise power for white noise and a given signal waveform. For any filter, an efficiency factor may be defined as

$$n = \sqrt{\frac{\text{peak output signal power}}{\text{output noise power}}} \qquad (3.59)$$

for a given ratio of input signal energy to noise power density. In the band-dividing demodulator, the input signal is a square pulse of width τ_s and unit amplitude. It is convenient to define the filter's zero-frequency response as $|H(jo)|^2 = 1$. The output noise power from the filter is given by

$$\overline{n^2}_{out} = \frac{N_o}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)|^2 \, d\omega$$

$$= \frac{N_o}{2\pi} B_n \qquad (3.60)$$

where B_n is the usual 2-sided noise bandwidth of the filter in rad/sec. Setting $N_o/2\pi$ arbitrarily equal to 1, the efficiency factor becomes

$$n = \frac{\int_o^{T_s} n(t) \, dt}{\sqrt{B_n}} \quad , \qquad (3.61)$$

with the additional restriction that

$$\left| \int_{-\infty}^{\infty} h(t) \, dt \right| = |H(jo)| = 1 \, . \qquad (3.62)$$

The matched filter for a square pulse has a transfer function;

$$H_m(j\omega) = \frac{\sin \omega\tau_s/2}{\omega\tau_s/2} \quad , \qquad (3.63)$$

a noise bandwidth of

$$B_{n,m} = \frac{2\pi}{T_s} \qquad (3.64)$$

and a peak output signal amplitude of 1. Its efficiency factor is thus

$$n_m = \sqrt{\frac{T_s}{2\pi}} = 0.4\sqrt{T_s} \quad . \tag{3.65}$$

The single-pole filter has an equivalent lowpass transfer function

$$H_\alpha(j\omega) = \frac{\alpha}{j\omega + \alpha} \quad , \tag{3.66}$$

a noise bandwidth of

$$B_{n,\alpha} = \pi\alpha \quad , \tag{3.67}$$

and an impulse response $h_\alpha(t) = \alpha e^{-\alpha t}$.

Its efficiency factor is,

$$n_\alpha = \frac{1 - e^{-\alpha T_s}}{\sqrt{\pi\alpha}} \quad . \tag{3.68}$$

The efficiency is maximum when $\alpha = 1.25/T_s$, and is given by

$$n_\alpha \Big|_{max} = 0.36\sqrt{T_s} \quad . \tag{3.69}$$

The single-pole filter is therefore worse than the matched filter by

$20 \log(0.4/0.36) = 0.9$ db.

The 3-db half-bandwidth of the most efficient single-pole filter is

$$f_{co} = \frac{\alpha}{2\pi} = 0.20\, f_s \tag{3.70}$$

where f_s is the sample rate. It is obvious that even under steady-state

conditions a spacing of these filters by f_s (the sample rate) will lead to "holes" between filters where the signal response is several db down. There are two solutions to this problem. One is to reduce the spacing between filters and add more of them. The other is to increase the filter bandwidths, accepting some further sacrifice in efficiency. A factor of 2 increase in bandwidth will reduce the efficiency by 1.4 db. (The reduction in efficiency is less than 3 db because the noise power doubles and the signal output also increases somewhat.) The solution of adding more filters appears to offer less performance degradation. The probability of error in selecting a filter goes up roughly as the number of filters but a fraction of a db increase in input SNR will compensate for a 2 or 3 times increase in the number of filters. Barring practical considerations of complexity, it appears that a filter spacing of $f_s/2$ or $f_s/3$ is best for single-pole filters. The estimated frequency will then be quantized to finer steps and errors caused by inter-channel interference will be restricted largely to 1/2 or 1/3 of f_s instead of f_s.

It might be well to point out that even matched filters spaced by f_s will be 4 db down at the overlap point, and closer spacing of these filters is also necessary. The spacing of f_s suggested by Akima is based on the argument that the bandwidth of a pulse τ_s long is <u>roughly</u> $\pm f_s = 1/\tau_s$. The question of the optimum filter spacing is unresolved and cannot be obtained from Akima's model.

REFERENCES - CHAPTER III

1. Baghdady, E. J., Lectures on Communication System Theory, McGraw-Hill Book Co., New York, 1961; Chapter 19.

2. _____, "Theoretical Comparison of Exponent Demodulation by Phase-Lock and Frequency-Compressive Feedback Techniques," 1964 IEEE International Convention Record, Part 7.

3. Lehan, F. W. and Parks, R. J., "Optimum Demodulation," 1953 IRE National Convention Record, Part 8, pp. 101-103.

4. Akima, H., "Theoretical Studies on Signal-to-Noise Characteristics of an FM System," IEEE Trans. on Space Electronics and Telemetry, Vol. SET-9, pp. 101-108, December 1963.

5. Battail, G., Sur le Seuil de Reception en Modulation de Frequence Annales des Telecommunications, t. 19, nos. 1-2, 1964.

6. Reiger, S., "Error Rates in Data Transmission," Proc. IRE, Vol. 46, pp. 919-920, May 1958.

IV. EVALUATION OF CERTAIN PUBLISHED ANALYSES

4.1 Quasi-Linear Model of PLL Demodulator

In two recent papers[1,2] on phase-lock reception, Develet has advocated a "quasi-linear" model of phase-locked loops for obtaining "a simple analytical threshold criterion." The proposed model and the treatment and results presented deserve careful assessment and criticism.

We start with the first paper, which considers "reception of an unmodulated sinusoid embedded in additive white gaussian noise." The discussion starts with a "simplified block diagram of general phase-lock receiver with no signal modulation imposed" in which the input noise is represented only by its quadrature component denoted Y(t). That this model and representation are questionable may be immediately seen from Eq. (3.7) with $\psi(t) \equiv 0$, namely

$$\frac{1}{\mu\alpha}\dot{\phi}_{osc}(t) = \left\{-\left[1+\frac{1}{E_s}x_{c,if}(t)\right]\sin\phi_{osc}(t) + \frac{1}{E_s}x_{q,if}(t)\cos\phi_{osc}(t)\right\} \otimes h_{lp}(t) \qquad (4.1)$$

subject to the condition $|\epsilon(t)| \equiv |\phi_{osc}(t)| \leq \pi/2$. The model suggested by this equation is shown in Fig. 8. The corresponding model presented by Develet in his Fig. 2 differs from the correct one based on Eq. (4.1) by the unjustifiable omission of the block marked cos ( ) in our own Fig. 8.

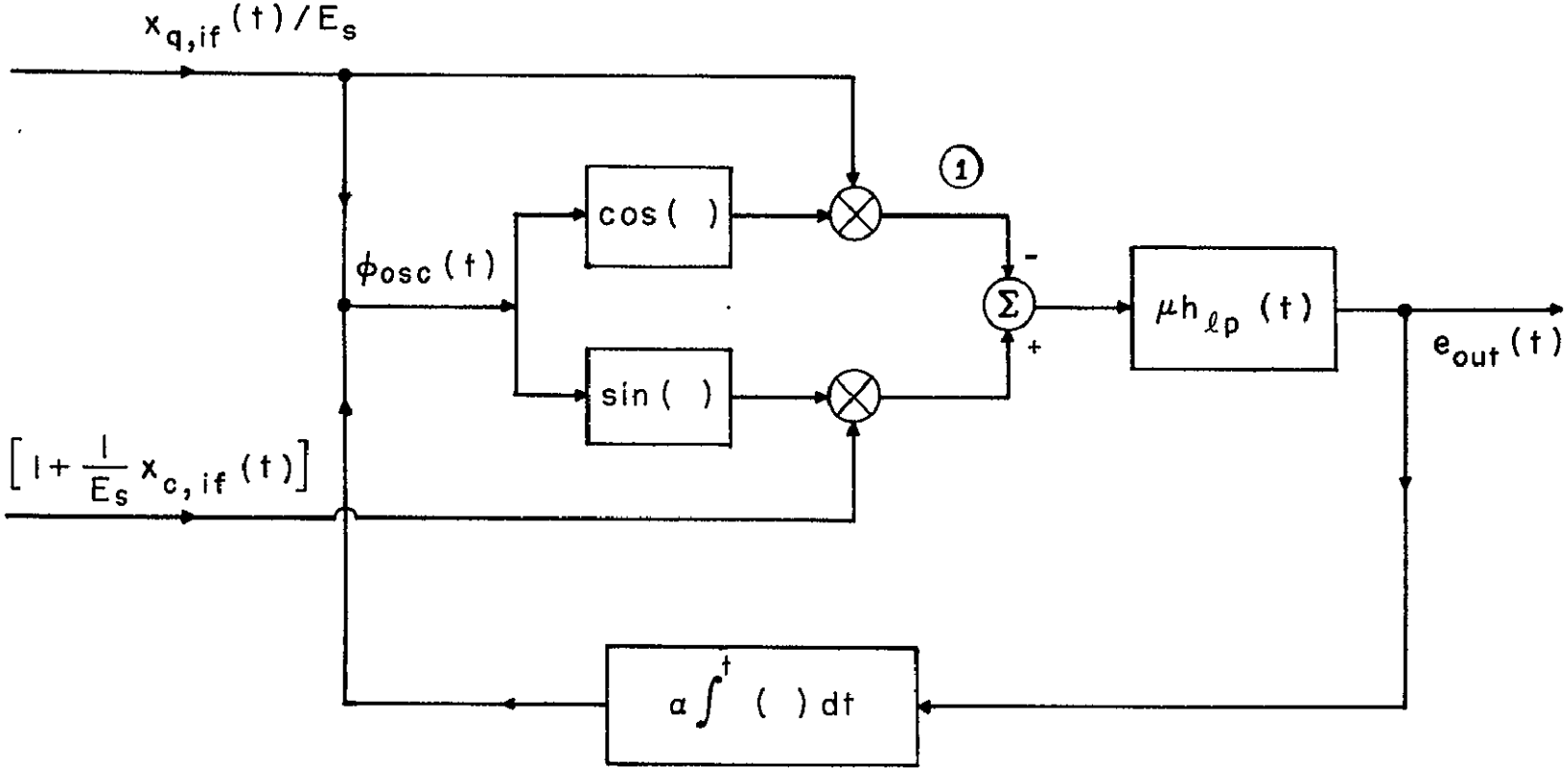


Fig. 8 Model of PLL for demodulation of unmodulated signal plus noise, subject to the condition $|\phi_{osc}(t) \leq \pi/2$.

two multipliers within the loop, and the $\left[1 + \frac{1}{E_s} x_{c,if}(t)\right]$ component of the input. These omissions are objectionable because:

a) The elimination of cos x imposes a more stringent restriction on $|x|$ than does the linearization of sin x – the first rejected term in the approximation $\cos x \approx 1$ is $x^2/2$ as opposed to $x^3/6$ for $\sin x \approx x$. Therefore, omitting cos x while at the same time retaining sin x unlinearized is indefensible and not allowable.

b) The omission of the input component $1 + \frac{1}{E_s} x_{c,if}(t)$ amounts to approximating this component by unity. This approximation is allowable provided that $\frac{1}{E_s} |x_{c,if}(t)| \ll 1$ almost all of the time and this is practically considered to hold only if

$$(S/N)_{if} \geq 5 \text{ (or 7 db)}$$

With regard to (b), note that even if one goes out of the way and sets $\sin \phi_{osc}(t) \approx \phi_{osc}(t)$ and assumes that the closed-loop noise bandwidth of the resulting linear (not quasi-linear) model is much smaller than the i-f bandwidth (which is usually justifiable for the scope of Ref. 1, but not in general for Ref. 2), the requirement for omitting $1 + \frac{1}{E_s} x_{c,if}(t)$ from the model is only reduced to

$$(S/N)_{c\ell} \geq 5 \text{ (or 7 db)}$$

in which $(S/N)_{c\ell}$ is the ratio of mean-square signal to mean-square noise as filtered by the bandpass analog of the closed-loop low-pass equivalent.

Since the object of Develet's treatment is the analysis of a threshold which according to his computations should occur down at a S/N within the closed-loop bandwidth of 1.34 db, the indicated omission is completely unallowable.

The quasi-linearization and analysis of the model considered by Develet in his Fig. 2 are also questionable. The quasi-linearization procedure employed "essentially determines the average gain of the nonlinear device [sin ( )] under the expected operating conditions." The objections against this are:

a) The replacement of sin ( ) by an average gain is subject to restrictions on the allowable values of the argument of the sine function which are neither stated nor heeded by Develet.

b) This replacement is definitely invalid for a phase-locked loop unless $\phi_{osc}(t)$ (or $\epsilon(t)$ in Develet's notation) is restricted in magnitude, because it cannot account for the break of lock that results whenever the condition

$$|\phi_{osc}(t)| \leq \pi/2$$

is violated. The condition for lock can be expected to be satisfied almost all of the time as long as

$$\sigma^2 \equiv \phi^2_{osc}(t) \leq \pi^2/64 \approx 1/6.4$$

where we have allowed a gaussian-like character for $\phi_{osc}(t)$ with a crest factor of 4. This shows that Develet's model breaks down for $\sigma > \pi/8 \approx 0.4$ rad. Account of this breakdown of the model must be taken in defining the limits

of the integration in Develet's Eq. (2). But since the results and curves cannot be used for $\sigma > 0.4$ rad, the necessary correction for the limits of the integration will not have a substantial effect. According to the curves of Fig. 5, the "loop signal to noise ratio" for $\sigma = 0.4$ is ≈ 5.2 db for the "quasi-linear approximation" and ≈ 5 db for the "completely linear approximation." These are then the thresholds that can be read from Develet's computations. (A more complete account of this threshold is presented in Ref. 3.)

Philosophically, the concept of replacing "instantaneous" effects by the effect on the average is disquieting in the present application because the breakdown of performance is caused by disruptions resulting from excesses by the instantaneous values of the phase error that cause the system to be either in lock, and hence describable by the model used, or out of lock, and hence not describable by the model used.

It also appears that approximation of a substantially nonlinear input-output characteristic by its average gain over the distribution of an input random process is a great oversimplification in the computation of the effects of this nonlinearity upon the input process, or upon the sum of this and another process.

It is of interest to point out that the curve marked "quasi-linear approximation" in Develet's Fig. 4 is incorrect. The reason for this is that the B_n in the left-hand member of Develet's Eq. (7) is the noise

bandwidth of the closed-loop system of his Fig. 3. This B_n is therefore function of σ. In plotting his curve, Develet has ignored this dependence on the basis of the statement that holding B_n constant "... can theoretically be accomplished by varying the gain of F(s) in opposition to exp $(-\sigma^2/2)$." Varying the gain of F(s) in opposition to exp $(-\sigma^2/2)$ means precisely replacing F(s) by F(s) exp $(\sigma^2/2)$. If this is indeed done, the open-loop transfer function of Develet's model becomes identical with that of what he calls the "completely linear" model, and the "quasi-linear" model therefore becomes "the completely linear" model. Ignoring the dependence of B_n upon σ is therefore unjustified and the "quasi-linear approximation" curve in Fig. 4 does not correctly portray the performance of the model of Fig. 3.

To sum up the above critique of Ref. 1:

a) The model of Fig. 2 is an incorrect representation of the "general phase-lock receiver" described in Fig. 1.

b) The model of Fig. 2 cannot be used for analysing the performance of phase-lock receivers at or near their threshold of performance breakdown.

c) The quasi-linear model employed omits vital effects of nonlinear behavior at or near the threshold of performance breakdown.

d) Even when proper corrections are made throughout for model, expressions and curves, the validity of the results and curves must be restricted to values of rms

loop error σ less than approximately 0.4 rad, because above this value of σ a quasi-gaussian phase error distribution would indicate that unlock becomes likely during an intolerable fraction of the time.

In his second paper,[2] entitled, "A Threshold Criterion for Phase-Lock Demodulation," Develet considers phase or frequency modulation by "Gaussian signals corrupted by additive white Gaussian noise." At the outset, a restrictive assumption is made and used in modeling the system which restricts the validity of the model only to the case in which the predetection bandwidth is much larger than the phase-lock loop bandwidth. This assumption makes it allowable to consider the noise in the low-frequency output of the product demodulator to have a gaussian character and a correlation function (or spectral distribution) identical with that of the cophasal and quadrature components of the input noise. (The results of the analysis are applied extensively in a later portion of the paper to cases in which this basic assumption does not hold at all. Specifically, note that curves are plotted and discussed for rms deviation ratios of less than 10 in Figs. 4 and 5.) As in the first paper, discussed above, the fact that break of lock ultimately limits the usefulness of the output of a phase-locked demodulator, and hence incurs the ultimate threshold, is completely ignored in this second paper. Thus, the basic model and analysis used in this paper are inadequate and invalid because:

a) The error signal $\epsilon(t)$ is not restricted between $\pm\pi/2$ in order to avoid unlock.

b) Unlock introduces severe disturbances that cannot be characterized by the model of injected noise used in the paper.

c) The replacement of the nonlinear characteristic, sin ( ), by its average derivative over a gaussian distribution is improper because the error signal $\epsilon(t)$ is not in general gaussian, and because important by-products of the nonlinearity are thereby excluded whose instantaneous (rather than average) fluctuations have a critical effect upon the threshold performance of a phase-locked loop.

d) It is not realized in the formulation of the model that in order to keep nonlinear distortion in the absence of noise below tolerable limits, the magnitude of the signal component of $\epsilon(t)$ must be restricted to values well below unity.

4.2 Develet's Least Squares PLL Optimization

Develet[2] has also carried out a least squares optimization analysis of the PLL demodulator. There is doubt about Develet's results for two reasons:

1. He uses an unrealistically large loop error, $\sigma^2 = 1$.
2. He uses the quasi-linear equivalent gain model which seems to be of little value because it considers only one of the effects of the nonlinearity.

These doubts justify repeating the analysis with a considerably smaller value of σ^2 (0.1) and without quasi-linearization. However, the results of our analysis (Chapter V of this report) agree with Develet on the one point most important to this project: there is about a 3 db difference in threshold between an optimized second-order system and the optimum system at a modulation index of 10.

4.3 Sanneman and Rowbotham Paper

The digital simulation of a second-order PLL by Sanneman and Rowbotham[4] has produced an interesting result which was not fully developed in their paper. They used a sampled-data system to simulate the PLL; the sampling rate being chosen large enough to ensure correspondence with the continuous system. The computerized loop was supplied with a simulated carrier plus noise input and allowed to run (starting with prescribed initial conditions) until a break of lock occurred. A large number of runs were taken for each set of conditions so that an average time to unlock could be computed for various carrier-to-noise ratios and initial conditions.

The authors make an issue of the fact that the probability of unlocking in τ seconds can be described by the empirical law

$$P(\omega_n \tau) = 1 - e^{-\omega_n \tau / \omega_n \bar{\tau}} , \qquad (4.2)$$

where $\bar{\tau}$ is the average time to unlock. This function is simply the probability of one or more Poisson events in an interval $\omega_n \tau$ and would be obtained for

a model where the probability of unlock is constant in any given time interval. In other words, their data supports the hypothesis that unlock events occur individually and collectively at random.

Of greater interest is the plot of average time to unlock versus phase error given in their Fig. 12. The abcissa of this plot is directly the rms phase error and the ordinate is the log of the normalized time to unlock. Viterbi[3] has obtained an expression for the mean time to unlock for the filter-less PLL which takes the form

$$B_n \bar{\tau} \approx \frac{\pi}{4} \exp\left[\frac{2.0}{\sigma_\phi^2}\right] \tag{4.3}$$

where B_n is the noise bandwidth of the loop. A plot of $\log B_n \bar{\tau}$ versus $1/\sigma_\phi^2$ yields a straight line. This result immediately suggests a re-plotting of S and R's Fig. 12 with an abcissa of $1/\sigma_\phi^2$. Such a plot is shown in our Fig. 9. The data now yields a straight line. For both initial conditions equal to zero, S and R's data is described by

$$\omega_n \bar{\tau} = 2 \exp\left[\frac{1.60}{\sigma_\phi^2}\right] \tag{4.4}$$

Equation (4.4) therefore appears to be an adequate empirical expression for the mean time to unlock in the second-order PLL, a result which Viterbi was unable to obtain theoretically. The use of Eq. (4.4) is further justified by the fact that it is very similar to the theoretical result in Eq. (4.3) for the filterless loop.

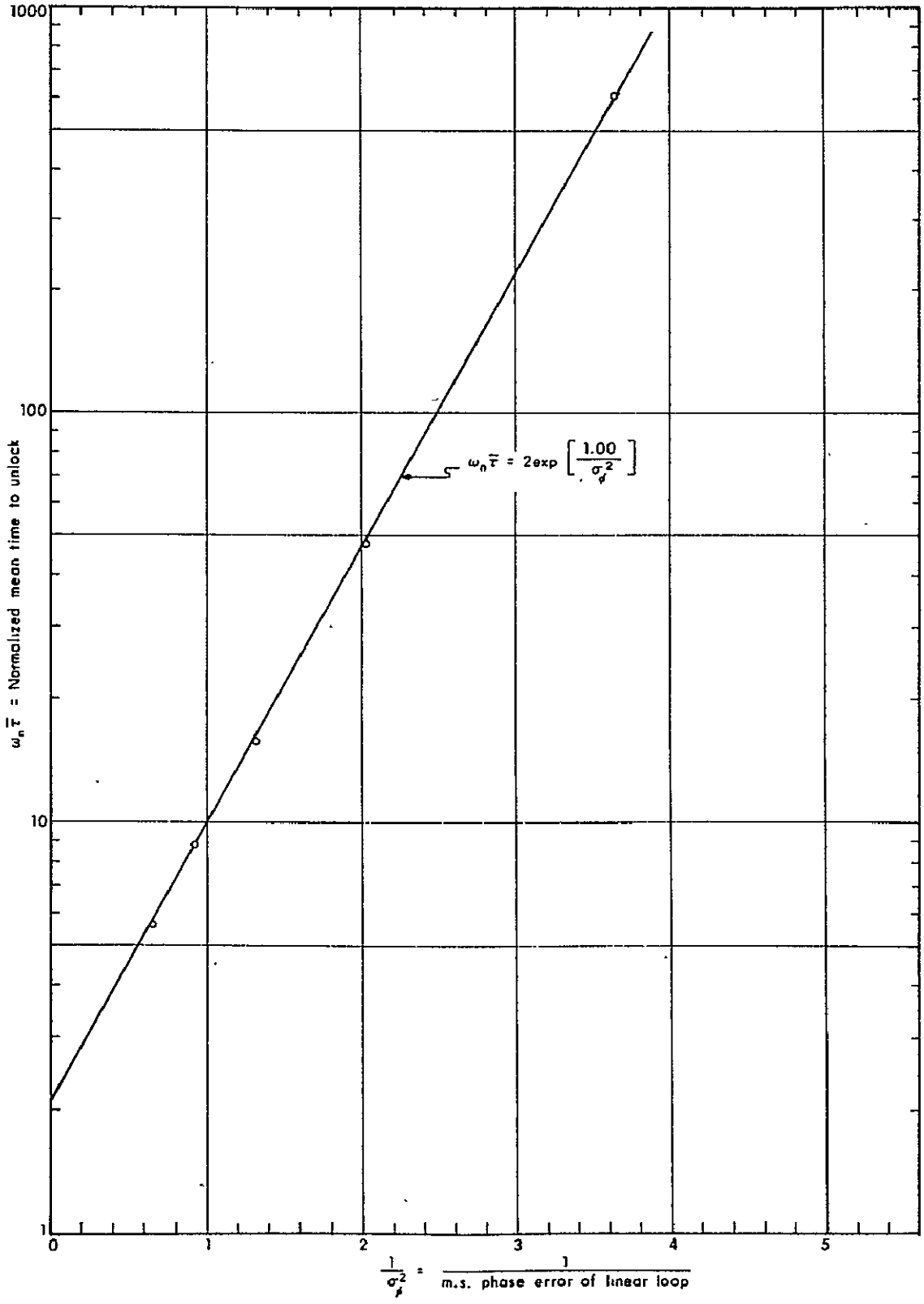


Fig. 9 Replot of data in Sanneman and Rowbotham's Fig. 12.

It is now possible to define a threshold value for σ_ϕ^2 in terms of a prescribed number of unlocks per second. For example, a second-order loop having a 16 kc natural frequency ($\omega_n = 105$) will have an average of 1 unlock per second at a mean-squared phase error of $\sigma_\phi^2 = 0.15\ \text{rad}^2$. For $\sigma_\phi^2 = 1$ as proposed by Develet[1,2] and others the corresponding unlock rate is 10,000 per second for $\omega_n/2\pi = 16$ kc.

These results are for an unmodulated carrier plus noise and zero steady-state phase error. If the mean-squared phase error includes a random modulation-tracking error the unlock rate will certainly be no less than that given by Eq. (4.4). Therefore, for audio-type loop bandwidths and a threshold defined by an unlock rate of about 1. second the threshold occurs at a mean-square phase error of $0.15\ \text{rad}^2$. The unlock rate is much more sensitive to the phase error than it is to the loop bandwidth, therefore one need not be extremely precise in specifying the loop bandwidth or the exact unlock rate. For example, a 10 times decrease in unlock rate corresponds to $\sigma_\phi^2 = 0.123$ instead of 0.15 or about 0.9 db difference in SNR measured in the loop bandwidth.

On the basis of Sanneman and Rowbotham's results, it therefore appears that an appropriate threshold value of total mean-squared phase error in the second-order PLL is $0.15\ \text{rad}^2$. At this level of error, the linearized model of the PLL is reasonably accurate. The average gain of

the multiplier, as computed by Develet[1], is 0.93 instead of 1. The threshold criteria of Develet[1,2] and Van Trees[5] which are based upon a point of divergence of mean-square phase error in an approximate model yield unreasonably high rates of unlock. The divergence of phase error (or gross unlock in some sense) is a phenomenon associated with the particular model being used. This phenomenon of the model is not an adequate description of the random breaks of lock which are observed to occur near the threshold.

REFERENCES - CHAPTER IV

1. Develet, J. A., "An Analytic Approximation of Phase-Lock Receiver Threshold," IEEE Trans. PGSET, 9, No. 1, pp. 9-11, March 1963.

2. _____, "A Threshold Criterion for Phase-Lock Demodulation," Proc. IEEE, pp. 349-356, February 1963.

3. Viterbi, A. J., PLL Dynamics in the Presence of Noise by Fokker-Planck Techniques, Proc. IEEE, December 1963.

4. Sanneman, R. W. and Rowbotham, J. R., "Unlock Characteristics of the Optimum Type II Phase-Locked Loop," IEEE Trans. PTGANE, Vol. ANE-11, No. 1, pp. 15-24, March 1964.

5. Van Trees, H. L., "An Introduction to Feedback Demodulation," M.I.T. Lincoln Lab. Report, Contract AF19628500, AD 416 586.

V. HIGHER-ORDER PHASE-LOCKED LOOP PERFORMANCE

5.1 Introduction

The use of a sinusoidal FM message for the purpose of comparing different PLL structures is supported by the following reasons:

a) The analytical evaluation of system performance is rather straightforward which permits one to keep a close tag on the analysis assumptions and limitations, common causes of serious trouble in the study of PLL systems.

b) A comparison with existing conventional demodulator analyses is possible, though the PLL threshold analysis must demand both message and noise to be present at the same time due to the signal tracking vs. noise rejection compromise of the system, a condition which may be somewhat relaxed in conventional demodulation.

c) The demand to match the laboratory experiments to the theoretical analysis can be met more easily.

Meanwhile, the use of higher order loops is perhaps motivated by the work of Jaffe and Rechtin* where a second order loop was found to be optimum for a frequency step message and a third order loop for a frequency ramp message. They considered a linear, noisy (additive noise with flat power spectral density), time invariant model and a minimum total (message

---

* R. Jaffe and E. Rechtin, "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance over a Wide Range of Input Signal and Noise Levels", IRE Trans. on Info. Theory, March, 1955.

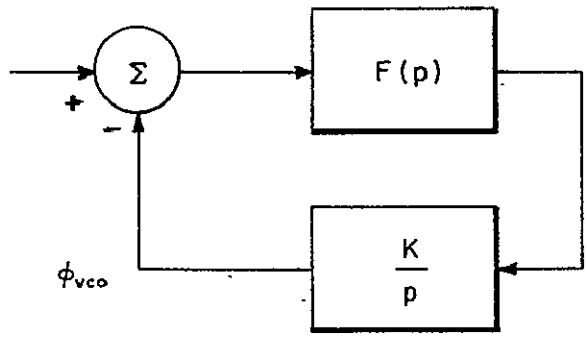
distortion plus noise) mean-square error as the optimization criteria. In any case, the only hint we want to draw is that the third order loop can be expected to do a good job on the small angle region of the sinusoidal FM signal where $\sin x \approx x$ allows a ramp approximation, while there is no equivalent motivation for a frequency step characterization and, consequently, a second order loop.

5.2 Phase Error and Threshold Considerations

The linear, time invariant phase model of a PLL is assumed as a starting point. This inherently assumes an approximation to the problem, since the loop phase noise and the additive input noise will show some statistical peaks that exceed the linearity and time invariancy conditions with a finite probability such that the model in question is not valid always. However, such behavior is assumed to have a small frequency of occurrence so that the model may be assumed as valid "almost all the time." The threshold phenomena essentially represent the breakdown of this model and are thus characterized by the occurrence of the statistical peaks with a frequency (probability measure) larger than allowable.

Consider the model of interest as shown in Fig. 11, where the input phase ϕ_{in} is the message phase ϕ_s plus the normalized quadrature noise ϕ_n, and ϕ_{vco}, K and F(p) respectively represent the VCO output phase, loop gain and loop filter. The loop phase error in tracking the message ϕ_s is given by

$$\phi_{in} = \phi_s + \phi_n$$



$$H(s) \triangleq \frac{\phi_{vco}(s)}{\phi_{in}(s)} = \frac{KF(s)}{s + KF(s)}$$

Fig. 10 Linear, time invariant phase model.

$$\phi_e(s) = \phi_s(s) - \phi_{vco}(s) = \phi_s(s) - H(s)[\phi_s(s) + \phi_n(s)]$$

$$= [1 - H(s)]\phi_s(s) - H(s)\phi_n(s) \qquad (5.1)$$

where the first term is a message distortion term characterizing the inability of the loop to exactly reproduce the message and being denoted as the (linear) distortion error $\phi_{e,d}$, while the second term is a noise term representing the VCO phase noise caused by the additive input noise and being denoted as the noise error $\phi_{e,n}$

In the case of a sinusoidal FM message having a modulation index δ and a modulation frequency ω_m rad/sec, i.e.,

$$\phi_s(t) = \delta \sin\omega_m t\,, \qquad (5.2)$$

the distortion error is sinusoidal with an amplitude given by

$$|\phi_{e,d}| = \delta\,|1 - H(j\omega_m)| \text{ rad} \qquad (5.3)$$

so that $|1 - H(j\omega_m)|$ essentially represents a compression factor for the modulation index. Meanwhile, if the noise has a normalized flat power spectral density of $\Phi/2S$ (rad/cps) and zero mean, the rms phase noise error is given by

$$\phi_{e,n_{(rms)}} = \left[\frac{\Phi}{2S}\int_{-\infty}^{\infty}|H(j\omega)|^2\,\frac{d\omega}{2\pi}\right]^{1/2}$$

$$= \left(\frac{\Phi B_n}{2S}\right)^{1/2} \text{ rad} \qquad (5.4)$$

where $B_n = \int_{-\infty}^{\infty}|H(j2\pi f)|^2\,df$ is the equivalent phase noise bandwidth of the

system in cps. If the statistics of this phase noise are known, then the probability of exceeding F times the rms value can be evaluated for any F and a statistical peak is thus defined when a sufficiently small probability is achieved.

On this basis, the total peak phase error may be taken as

$$\phi_{e,p} \simeq |\phi_{e,d}| + F\phi_{e,n(rms)} \tag{5.5}$$

where the "crest factor" F may be readapted further to acknowledge the fact that the distortion and noise error peaks do not occur necessarily at the same time instant. Finally, if ϕ_c characterizes the critical phase error beyond which the model is assumed to break down, then operation above the threshold "almost all the time" is characterized by $\phi_{e,p} \leq \phi_c$.

The choice of ϕ_c is linked to the crest factor F. If the noise statistics are assumed gaussian, then the use of F = 1, 2, 3 implies that the "peak" value $F\phi_{e,n(rms)}$ will be exceeded only 33%, 3%, 1% of the time, respectively. On this basis, the use of F = 3 and $\phi_c = \pi/2$ seems suitable. In reality, the noise is not gaussian near threshold and the PLL model is nonlinear near $\pi/2$ and time variant near threshold. Still, these parameters are but a performance criterion and should provide an adequate guidance to the threshold phenomenon. A proof of this statement is perhaps the fact that $3\phi_{rms} < \pi/2$ (neglect distortion temporarily) implies $\phi_{rms} < \pi/6$ which is a common upper bound in the linear analysis of PLL systems. The inclusion of the distortion term will yield a somewhat smaller rms

noise error but not by a considerable amount if the loop is operating in an optimized behavior, as shown in the sections that follow.

5.3 The Second Order Loop

The second order loop is characterized by a loop filter of the form

$$F(s) = \frac{1 + \tau_1 s}{1 + \tau_2 s} \tag{5.6}$$

which results in a noise bandwidth given by

$$B_n = \frac{K}{2} \frac{1 + \frac{K\tau_1^2}{\tau_2}}{1 + K\tau_1} \quad \text{cps}$$

$$\approx \frac{1}{2}\left(\frac{1}{\tau_1} + \frac{K\tau_1}{\tau_2}\right) \quad \text{for } K\tau_1 \gg 1 \tag{5.7}$$

At this point it is useful to notice that this bandwidth uniquely specifies the noise error for a given input noise density Φ/S (and noise statistics) so that several second order loops having different design parameters K, τ_1, τ_2 but a similar B_n will show the same noise rejection capability and their relative merits will only be determined by their message tracking (distortion error) capabilities. On this basis, it seems useful to change the set of design parameters (K, τ_1, τ_2) to a new set that has B_n as one of the parameters. A look at the approximate relation in Eq. (5.7) suggests the following transformation:

$$\frac{1}{\tau_1} = a B_n, \quad \frac{K\tau_1}{\tau_2} = (2 - a) B_n, \quad 0 < a < 2 \tag{5.8}$$

which yields the new set of parameters (K, B_n, a) where*

$$\tau_1 = \frac{1}{a B_n}, \quad \tau_2 = \frac{K}{a(2-a) B_n^2}, \quad \frac{K}{B_n} \gg a \tag{5.9}$$

The peak distortion error in terms of the new parameters is given, for the second order loop, by

$$|\phi_{e,d}| \approx \delta \left[\frac{\left(\frac{\omega_m}{K}\right)^2 + \frac{1}{a^2(2-a)^2}\left(\frac{\omega_m}{B_n}\right)^4}{1 + \frac{2-3a}{a^2(2-a)}\left(\frac{\omega_m}{B_n}\right)^2 + \frac{1}{a^2(2-a)^2}\left(\frac{\omega_m}{B_n}\right)^4} \right]^{1/2} \tag{5.10}$$

for $\frac{K}{B_n} \gg 2a$, and by noting that the maximum value of $a(2-a)$ is 1 and that K can be made very large, the first term in the root numerator may be dropped. Also, it seems reasonable to expect that either the first or the third term in the root denominator will predominate, except perhaps for a small range of ω_m/B_n values, which results in a compression factor of the form $\left(\frac{x}{1+x}\right)^{1/2}$ and $x \ll 1$ is desired if any noticeable compression is to occur, i.e.,

$$|\phi_{e,d}| \approx \frac{\delta \omega_m^2}{a(2-a) B_n^2} \tag{5.11}$$

---

* The exact noise bandwidth is actually given by $B_n/[1+(aB_n/K)]$ approximates B_n for $K/B_n \gg a$, a reasonable relation since $0 < a < 2$.

Thus, optimum performance occurs for $a = 1$ where the compression is maximum for a given noise error (i.e., B_n).

A more exact analysis can be done by plotting the compression factor with only the first term in the root numerator being neglected, i.e.,

$$|\phi_{e,d}| \approx \delta \left[1 + \frac{(2-a)(2-3a)}{\gamma^2} + \frac{a^2(2-a)^2}{\gamma^4} \right]^{-1/2} \qquad (5.12)$$

where $\gamma = \omega_m / B_n$ and the results are essentially shown in Fig. 11. For a constant γ, i.e., a given noise error, the interest is to operate at the largest possible abcissa in order to compress the modulation index as much as possible. Operation at $\gamma > 1$ is inadmissible since no essential compression occurs, while operation at very small γ's is excellent though the price is an extremely large noise bandwidth and noise error. In general, it is noted that $a = 1$ is optimum for $\gamma < 1$ where a compression occurs. Moreover, for $\gamma < 0.7$ the approximation of the abcissa by $\gamma - 2$ (at $a = 1$) can be found to show a negligible error. On this basis, our choice of Eq. (5.11) with $a = 1$ is truly an optimum behavior. Notice finally that the neglected factor containing the ω_m / K term will not alter the $\underline{a}$ choice but will only introduce a reduction in the assumed compression.

The total error after the first round of optimization (K large, $a = 1$) is thus given by

$$\phi_{e,p} = \frac{\delta \omega_m^2}{B_n^2} + F\left(\frac{\Phi B_n}{2S}\right)^{1/2} \qquad (5.13)$$

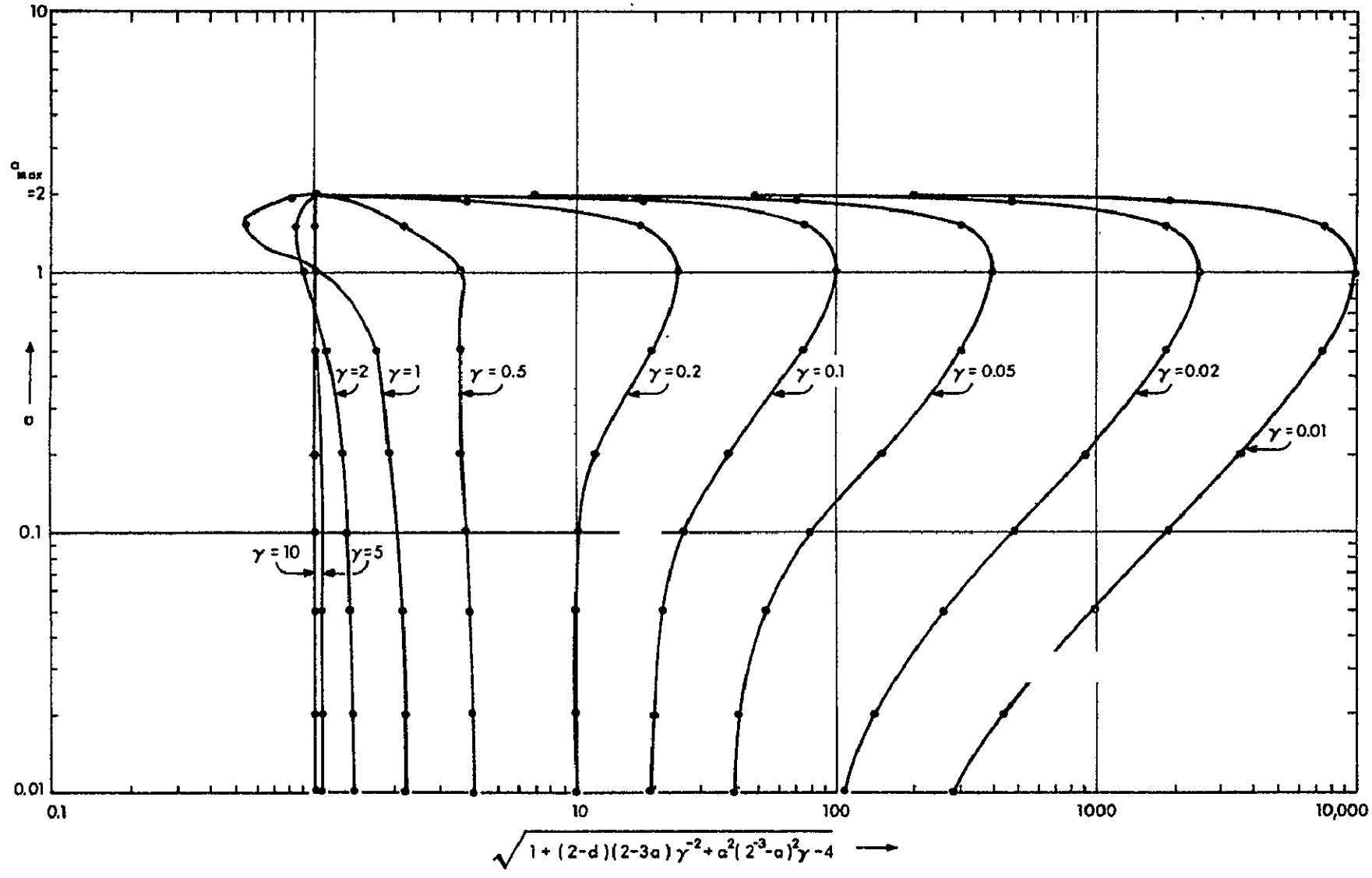


Fig. 11 Compression vs (a, γ).

and the final optimization is the choice of B_n for a minimum error, which can be found to be

$$B_{n_{opt}} = \left(\frac{4\delta\omega_m^2}{F}\right)^{2/5}\left(\frac{2S}{\Phi}\right)^{1/5} \text{ cps} \qquad (5.14)$$

Moreover, if this error is to be kept below ϕ_c "almost all the time," the threshold condition reads

$$\frac{S}{\Phi} \geq \frac{1.75F^2}{\phi_c^{5/2}}\delta^{1/2}\omega_m \qquad (5.15)$$

It is of interest to further analyze Eq. (5.13) with B_n given by Eq. (5.14). Direct substitution will prove the fact that the noise error term will be exactly 4 times larger than the distortion error term under this optimized behavior. This means that the total error can be considered as 5 times the distortion error (the other 4 contributions representing the bandwidth limitation introduced by noise) or as 5/4 times the noise error (the other 1/4 contribution representing the distortion limitation). By using the first characterization, the threshold condition reads

$$B_{n_{opt}} \geq \left(\frac{5\delta}{\phi_c}\right)^{1/2}\omega_m \text{ cps} \qquad (5.16)$$

while the second characterization yields a threshold condition of

$$(SNR)_{B_{n_{opt}}} \geq 0.78\left(\frac{F}{\phi_c}\right)^2 \qquad (5.17)$$

These two conditions respectively give the loop bandwidth required for tracking the message in the presence of noise and the loop SNR required for maintaining message lock in the presence of noise, i.e., the PLL threshold SNR.

If the intrinsic SNR is defined as

$$(SNR)_i = (SNR)_{2f_m}$$

then the intrinsic SNR at threshold is given by

$$(SNR)_i = (SNR)_{B_{n_{opt}}} \frac{B_{n_{opt}}}{2f_m}$$

From Eqs. (5.16) and (5.17) this ratio is

$$\text{db}\,(SNR)_i = 12\,\text{db} + \frac{10 \log \delta}{2} \tag{5.18}$$

for a crest factor $F = 3$ and $\phi_c = \pi/2$.

5.4 The Third Order Loop

The third order loop is characterized by a loop filter of the form

$$F(s) = \frac{1 + \tau_1 s + (\tau_3 s)^2}{1 + \tau_2 s + (\tau_4 s)^2} \tag{5.19}$$

which results in a noise bandwidth given by

$$B_n = \frac{K}{2} \cdot \frac{1 + \frac{K\tau_3^4}{\tau_4^2} \cdot \frac{\tau_1 + \frac{1}{K}}{\tau_3^2 + \frac{\tau_2}{K}} + \frac{\tau_1^2 - 2\tau_3^2}{\tau_3^2 + \frac{\tau_2}{K}}}{1 + K\tau_1 - \frac{\tau_4^2}{\tau_3^2 + \frac{\tau_2}{K}}} \text{ cps}$$

$$\approx \frac{K\tau_3^2}{2\tau_4^2} \cdot \frac{1 - \frac{K\tau_1\tau_3^2}{\tau_4^2} - \frac{\tau_1^2}{\tau_3^2}}{1 - \frac{K\tau_1\tau_3^2}{\tau_4^2} - \frac{\tau_3^2}{\tau_4^2}} \quad \text{for} \begin{cases} K\tau_1 \gg 1 \\ K\tau_3^2 \gg \tau_2 \end{cases} \tag{5.20}$$

The choice of a variable transformation in order to change the set $(K, \tau_1, \tau_2, \tau_3, \tau_4)$ to another one having B_n as a parameter is not obvious. However, an analogy with the second order loop suggests

$$\tau_1 = \frac{1}{aB_n} \quad , \quad \tau_3^2 = \frac{1}{bB_n^2} \quad , \quad \tau_4^2 = \frac{K}{cB_n^3} \tag{5.21}$$

where $\frac{K}{B_n} \gg a, \quad \frac{K}{B_n^2} \gg b\tau_2, \quad (a, b, c) > 0$

and $$a + 2 = \frac{2ab}{c} + \frac{c}{b} + \frac{b}{a} \tag{5.22}$$

On this basis, the new set of parameters is formed by K, B_n, τ_2 and any two of the three parameters (a, b, c), the other one being defined by Eq. (5.22).

The amplitude of the distortion error is now given, for the third order loop, by

$$|\phi_{e,d}| \approx \delta \left\{ \frac{1 - \frac{2c}{\gamma^3}\left(\frac{\omega_m}{K}\right) + \frac{c^2}{\gamma^6}\left[1 + (\tau_2\omega_m)^2\right]\left(\frac{\omega_m}{K}\right)^2}{1 + \frac{c}{\gamma^2}\left(\frac{c}{b^2} - \frac{2}{a}\right) + \frac{c^2}{\gamma^4}\left(\frac{1}{a^2} - \frac{2}{b}\right) + \frac{c^2}{\gamma^6}} \right\}^{1/2}$$

$$\text{for} \quad \frac{K}{B_n} \gg a, \quad \frac{K}{B_n^2} \gg b\tau_2 \tag{5.23}$$

where $\gamma = \omega_m/B_n$ as before. Again it seems that K and τ_2 can be chosen to neglect the last two terms in the root numerator, and that the last term in the root denominator may be found to predominate for an effective compression of the modulation index. In this case, Eq. (5.23) reads

$$|\phi_{e,d}| \approx \frac{\delta\omega_m^3}{cB_n^3} \tag{5.24}$$

and the interest is to select the largest possible value of c such that both Eq. (5.22) and the approximations that yielded Eq. (5.24) are satisfied. In general, it seems that $c \geq 1$ can be obtained, though an upper bound is difficult to establish since it implies component feasibility considerations.

On this basis, the total error is then given by

$$\phi_{e,p} = \frac{\delta\omega_m^3}{cB_n^3} + F\left(\frac{\Phi B_n}{2S}\right)^{1/2} \tag{5.25}$$

and the noise bandwidth optimization gives

$$B_{n_{opt}} = \left(\frac{6\,\delta\omega_m^3}{c\,F}\right)^{2/7} \left(\frac{2S}{\Phi}\right)^{1/7} \text{ cps} \qquad (5.26)$$

and the threshold condition of the phase error being kept below ϕ_c "almost all the time" reads

$$\frac{S}{\Phi} \geq \frac{1.29\,F^2}{c^{1/3}\phi_c^{7/3}}\,\delta^{1/3}\omega_m \qquad (5.27)$$

An analysis of the total error analogous to that effected for the second order loop results in the noise error term being 6 times larger than the distortion error term in the optimized behavior. By considering the total error as 7 times the distortion error, the threshold condition yields the following tracking requirement:

$$B_{n_{opt}} \geq \left(\frac{7\delta}{c\phi_c}\right)^{1/3}\omega_m \text{ cps} \qquad (5.28)$$

while by considering the total error as 7/6 times the noise error, the following loop threshold SNR is obtained:

$$(SNR)_{B_{n_{opt}}} \geq 0.68\left(\frac{F}{\phi_c}\right)^2 \qquad (5.29)$$

Corresponding to the second order loop the third order loop has an intrinsic SNR at threshold of

$$\text{db } (\text{SNR})_i = (\text{SNR})_{2f_m} = 11.1 \text{ db} + \frac{10 \log \delta}{3} \qquad (5.30)$$

for $F = 3$, $\phi_c = \pi/2$ and $c = 1$. If larger values of c can be obtained the SNR will be reduced by $10 \log c/3$ db. The threshold improvement over the second order loop is $0.9 + 1.7 \log \delta$ db.

5.5 Systems Comparison

A numerical comparison of the two system thresholds is dependent on the choice of the critical error ϕ_c and the parameter $\underline{c}$, as well as on the actual value of the modulation index. However, for $\phi_c \leq \pi/2$ and $c \geq 1$, the third order loop will always show a threshold improvement for reasonably large indices greater than one, the improvement being increased as $\underline{c}$ becomes larger.

There is another important advantage of the third order loop. A look at Eqs. (5.16) and (5.28) shows that reasonably large bandwidths may result to such extent that the fact that they may become comparable to the i-f bandwidths must be accounted for. The third order loop is shown to have a smaller loop bandwidth for a given message so that its B_{if}/B_n ratio will be larger.

Finally, it is of interest to notice that the two loop thresholds, i.e., Eqs. (5.17) and (5.29) are essentially similar.

5.6 Higher Order Loops

The results of the previous sections can be tentatively extended to higher order loops, i.e., higher order loop filters. In general, if the distortion error term can be written as inversely proportional to B_n^k, the noise error being always directly proportional to $B_n^{1/2}$, then the following relation will hold after the total error is minimized with respect to B_n:

$$(\text{noise error}) = \frac{2k+1}{2k} \ (\text{distortion error})$$

and the relevant effect will be the $B_{n_{opt}}$ and S/Φ dependence on $\delta^{\frac{1}{2k+1}}\omega_m$ such that both the system threshold and the required loop bandwidth are considerably reduced as k increases. The crucial issue is the capability of being able to express the distortion error in the desired form: on the positive side there are more parameters to play with while on the negative side there are more approximations involved (the analysis is already difficult for the third order loop).

5.7 Least-Square Optimization of PLL Performance

We shall next present the optimum performance attainable with a phase-locked loop and discuss implementation intended to achieve this performance.

5.7.1 Goals of Optimization Procedure

The PLL receiver of interest is shown in Fig. 12. The received signal r(t) is composed of the transmitted signal s(t) and noise n(t) which is white in the spectral region occupied by s(t). We make the following definitions:

$$s(t) = \sqrt{2}\, A \sin(\omega_c t + \delta m(t)) \tag{5.31}$$

$m(t)$ = Message, a random process with spectrum $S_m(\omega)$

δ = modulation index

$$v_{VCO}(t) = \sqrt{2}\, A \cos(\omega_c t + \delta \hat{m}_\ell(t)) \tag{5.32}$$

We decompose the noise as follows:

$$n(t) = \sqrt{2}\, A\, n_1(t) \sin(\omega_c t + \delta m(t)) + \sqrt{2}\, A\, n_2(t) \cos(\omega_c t + \delta m(t)) \tag{5.33}$$

$n_1(t)$ and $n_2(t)$ are low-pass equivalents of n(t) and have a flat spectrum of height $N_o/2A^2$ in the low-pass spectral region occupied by m(t). With these definitions we have (discarding double frequency terms which will not pass through the loop filter):

$$r(t) \times v_{VCO}(t) = A^2 [1+n_1(t)] \sin\delta [m(t) - \hat{m}_\ell(t)] + A^2 n_2(t) \cos\delta [m(t) - \hat{m}_\ell(t)] \tag{5.34}$$

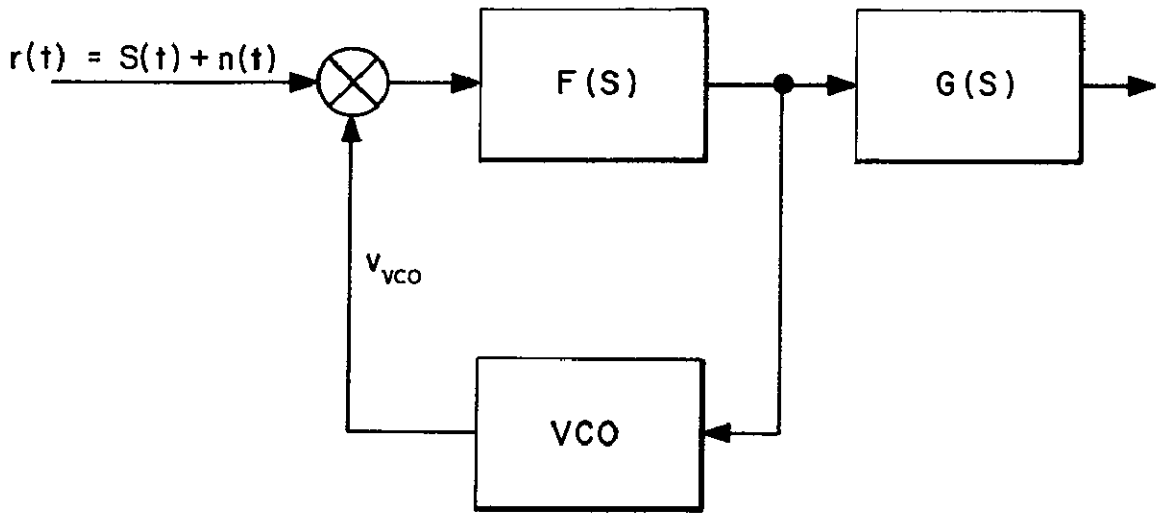


Fig. 12 Model of PLL.

Thus we may draw the non-linear, time-variant phase transfer model of Fig. 13. There are two effects of importance here. First, if the loop error $e_\ell(t)$ becomes larger than $\pi/2$ the loop will lose lock. Second, for small values of $e_\ell(t)$ the system is linear while for large values of $e_\ell(t)$ it is non-linear and time-variant. We can insure that operation will be linear most of the time by requiring that $VAR[\, e_\ell(t)\,] \leq \sigma_c^2$ where σ_c^2 is a suitably chosen constraint level. The magnitude of $VAR[\, e_\ell(t)\,]$ is also a measure of how often $e_\ell(t)$ will exceed $\pi/2$, causing the loop to lose lock. The phenomenon of threshold occurs when the loop loses lock with sufficient frequency to degrade performance. Threshold has no direct connection with the transition from linearity to non-linearity, but will occur for about the same values of $VAR[\, e_\ell(t)\,]$ which define this transition. Therefore, in the work that follows, we will postulate that threshold occurs for some loop error $VAR[\, e_\ell(t)] = \sigma_c^2$ and that operation of the loop is linear until threshold is reached. Thus, all following discussion will be based on the linear, time-invariant model of Fig. 14. Our goal will be to design F(s) so as to minimize $VAR[\, e_\ell(t)\,]$ and thus optimize the system with respect to threshold.

5.7.2 Linear Filter Optimization; Realizability

It is important to recognize that we must require F(s) to be realizable. In an ordinary filtering problem we may approximate the performance of an unrealizable filter by introducing delay between the input and output. However, in the phase-locked loop the filter output is required in order to form its output, so we cannot use this stratagem.

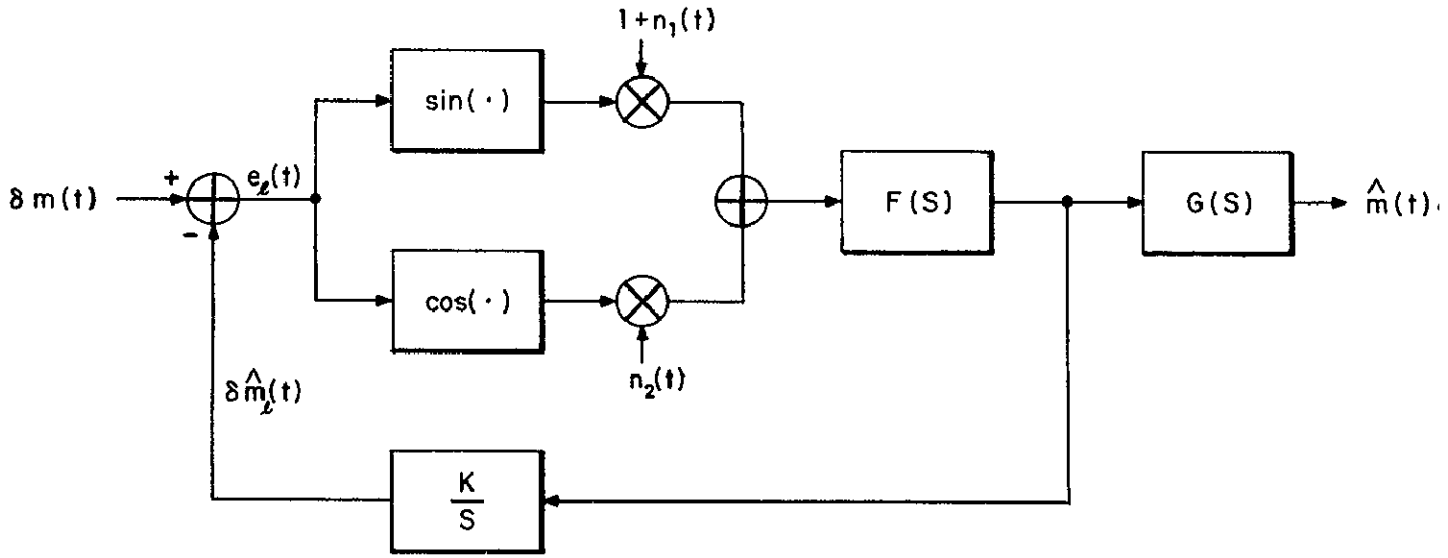


Fig. 13 Non-linear, time-varying phase transfer model of PLL.

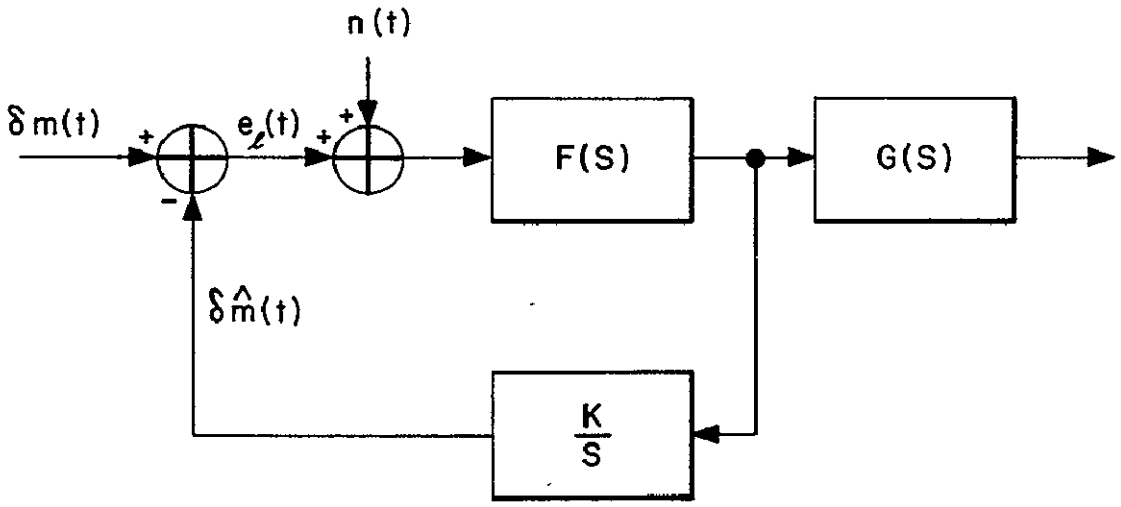


Fig. 14 Linear, time-invariant model of PLL.

The problem of the optimum linear filter has been solved in a convenient form by Yovits and Jackson[1]. Their results give the following expressions (in the white noise case) for the minimum error and for the filter which produces it:

$$\mathrm{VAR}\left[e_{\ell}(t) \right]_{opt} = \frac{N_o}{2\pi A^2} \int_o^{\infty} \log \left[1 + \frac{2A^2}{N_o} \delta^2 S_m(\omega) \right] d\omega \qquad (5.35)$$

and

$$\left| 1-H(j\omega) \right|^2 = \frac{1}{1 + \frac{2A^2}{N_o} \delta^2 S_m(\omega)} \qquad (5.36)$$

where we have defined the closed loop transfer function $H(j\omega)$ by

$$H(s) = \frac{KF(s)}{s + KF(s)} \qquad (5.37)$$

In order to calculate the loop error for non-optimum loop filters we must separate the two different causes of error. First there is a noise error $e_{\ell n}$ which is due to the presence of the noise source $n_2(t)$ in the loop. Second there is a distortion error $e_{\ell d}$ due to the fact that the loop transfer function is not identically unity over the spectrum of the message. These errors are given by

$$\mathrm{VAR}\left[e_{\ell n}(t) \right] = \frac{1}{\pi} \int_o^{\infty} \frac{N_o}{2A^2} \left| H(j\omega) \right|^2 d\omega \qquad (5.38)$$

and

$$\mathrm{VAR}\left[e_{\ell d}(t) \right] = \frac{1}{\pi} \int_o^{\infty} \delta^2 S_m(\omega) \left| 1-H(j\omega) \right|^2 d\omega \qquad (5.39)$$

Because the message and noise are independent, these errors are independent, and we may write

$$\mathrm{VAR}\,[e_{\ell d}(t)] = \mathrm{VAR}\,[e_{\ell n}(t)] + \mathrm{VAR}\,[e_{\ell d}(t)] \tag{5.40}$$

5.7.3 Performance of Optimum Realizable Loop for PM with Rectangular Message Spectrum *

We consider the message spectrum given by

$$(\omega) = \begin{cases} \pi/\alpha & |\omega| < \alpha \\ \\ 0 & \omega| > \alpha \end{cases} \tag{5.41}$$

We define the message bandwidth to be the bandwidth of a rectangular low-pass spectrum which has the same height at the origin as $S_m(\omega)$ and gives the same total power. Thus

$$B_{eq} = \frac{1}{S_m(o)} \int_{-\infty}^{\infty} S_m(\omega)\, d\omega \tag{5.42}$$

We will generally normalize $S_m(\omega)$ so that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S_m(\omega)\, d\omega = 1 \tag{5.43}$$

(Incidentally, this is the justification for calling δ the modulation index), hence

$$B_{eq} = 2\pi/S_m(\omega) \tag{5.44}$$

In the present case, of course, this reduces to

$$B_{eq} = 2\alpha \tag{5.45}$$

---

* This is the approach taken by Develet, Ref. 2.

Therefore the signal-to-noise ratio in the message bandwidth is given by

$$\Lambda = \frac{1}{\frac{N_o}{2A}\quad\frac{B_{eq}}{2\pi}} = \frac{2\pi A^2}{\alpha N_o} \tag{5.46}$$

Inserting (5.41) and (5.46) into Eqs.(5.35) and (5.36) we find

$$\mathrm{VAR}[\,e_\ell(t)]_{\mathrm{opt}} = \frac{1}{\Lambda}\ \log\ (1+\Lambda\,\delta^2) \tag{5.47}$$

and

$$|1-H(j\omega)|^2 = \begin{cases} \dfrac{1}{1+\Lambda\delta^2} & |\omega| < \alpha \\ 1 & |\omega| > \alpha \end{cases} \tag{5.48}$$

From Eq. (5.47) we can relate modulation index, message bandwidth SNR, and threshold as shown in Fig. 15. From Eq. (5.48) it is not clear how to construct the optimum loop filter; approximations to it will be discussed in section 5.7.5.

5.7.4 Optimum Realizable Loop for FM with R-C Message Spectrum

We take*

$$S_m(\omega) = \frac{1}{\omega^2(\omega^2+\alpha^2)} \tag{5.49}$$

and define

$$\gamma = \frac{2A^2}{N_o} \tag{5.50}$$

*This spectrum is not normalized in accordance with Eq. (5.43) because its integral is infinite. Therefore δ is not properly a modulation index.

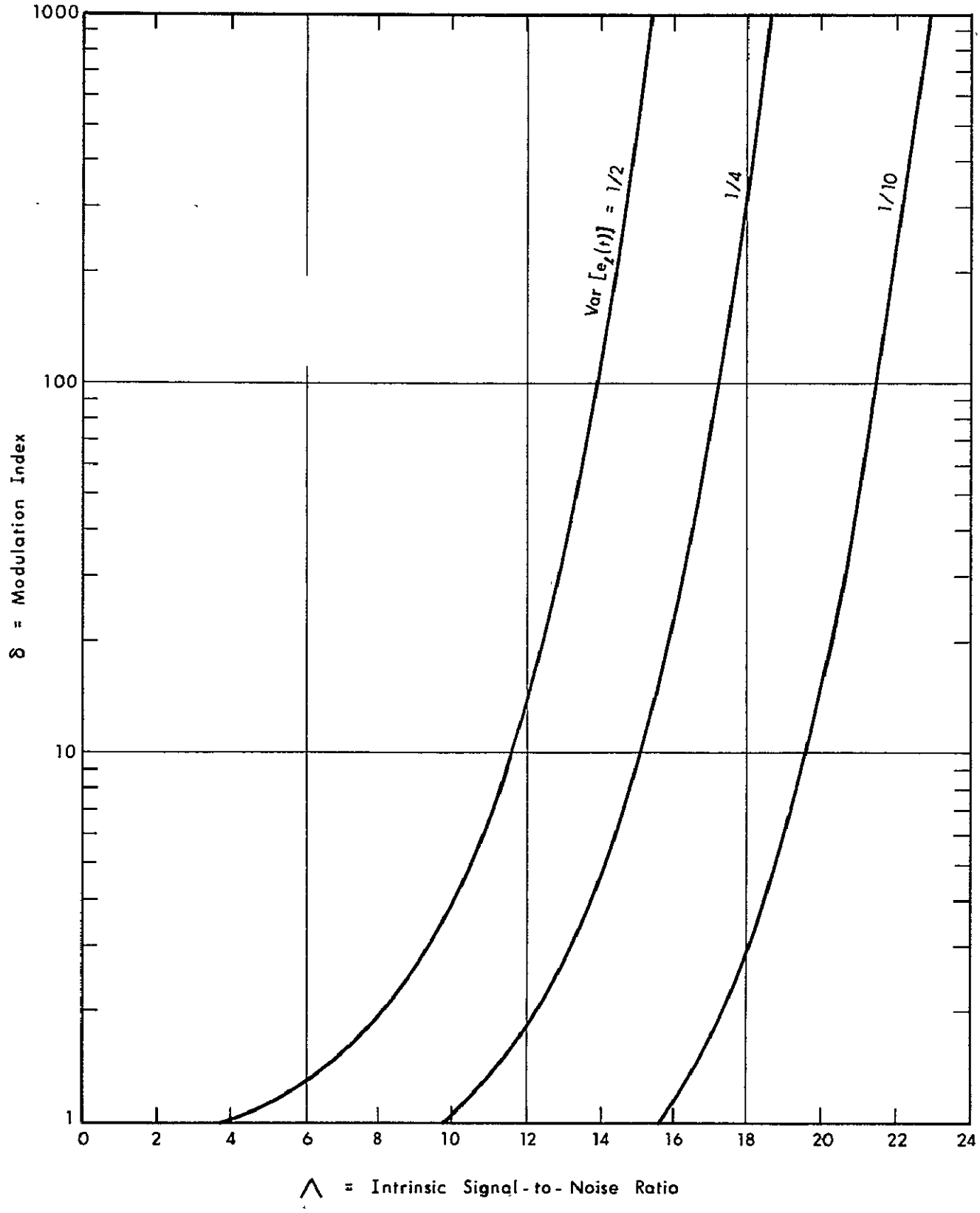


Fig. 15 Characteristics of optimum receiver.

Then from (5.36) we have

$$|1 - H(j\omega)|^2 = \frac{1}{1 + \gamma \delta^2 S_a(\omega)} = \frac{\omega^2(\omega^2 + \alpha^2)}{\omega^4 + \alpha^2\omega^2 + \gamma \delta^2} \tag{5.51}$$

Note that

$$\omega^4 + \alpha^2\omega^2 + \gamma \delta^2 = (\omega^2 + \rho_1)(\omega^2 + \rho_2) \tag{5.52}$$

where

$$\rho_1 = \frac{1}{2}(\alpha^2 + \sqrt{\alpha^4 - 4\gamma \delta^2}\,; \quad \rho_2 = \frac{1}{2}(\alpha^2 - \sqrt{\alpha^4 - 4\gamma \delta^2}) \tag{5.53}$$

So,

$$|1 - H(j\omega)|^2 = \frac{(-j\omega)(j\omega)(\alpha + j\omega)(\alpha - j\omega)}{(\rho_1^{1/2} + j\omega)(\rho_1^{1/2} - j\omega)(\rho_2^{1/2} + j\omega)(\rho_2^{1/2} - j\omega)}$$

and

$$1 - H(j\omega) = \frac{j\omega(\alpha + j\omega)}{(\rho_1^{1/2} + j\omega)(\rho_2^{1/2} + j\omega)} \tag{5.54}$$

Whence,

$$H(j\omega) = \frac{(\rho_1 \rho_2)^{1/2} + j\omega(\rho_1^{1/2} + \rho_2^{1/2} - \alpha)}{(\rho_1^{1/2} + j\omega)(\rho_2^{1/2} + j\omega)} \tag{5.55}$$

Note $\rho_1\rho_2 = \gamma \delta^2$ and furthermore

$$\rho_1^{1/2} + \rho_2^{1/2} = \sqrt{(\rho_1^{1/2} + \rho_2^{1/2})^2} = \sqrt{\rho_1 + \rho_2 + 2\rho_1^{1/2}\rho_2^{1/2}}$$

$$= \sqrt{\alpha^2 + 2(\gamma \delta^2)^{1/2}} \tag{5.56}$$

So

$$H(j\omega) = \frac{(\gamma\,\delta^2)^{1/2} + j\omega\,(\sqrt{\alpha^2 + 2(\gamma\,\delta^2)^{1/2}} - \alpha)}{(\rho_1^{1/2} + j\omega)\,(\rho_2^{1/2} + j\omega)} \tag{5.57}$$

The importance of this result is that it shows that the optimum loop for an FM system with an R-C message spectrum is a second order loop. That this is not true for PM with a rectangular message spectrum is evident from the form of Eq. (5.48).

5.7.5 Approximations to the Optimum Realizable Loop Filter for a Rectangular Message Spectrum

We found in section 5.7.3 that the form of the optimum loop filter is given by

$$|1 - H(j\omega)|^2 = \begin{cases} \dfrac{1}{1 + \Lambda\,\delta^2} & |\omega| < \alpha \\ 1 & |\omega| > \alpha \end{cases} \tag{5.58}$$

It is clear that to build a system having exactly this transfer function would require an infinite number of poles and zeros. However, we can certainly approximate it as closely as desired by using a system of high enough order. Because of the increase of complexity associated with higher order filters we would like to keep the order, n, as small as possible while meeting a desired performance criterion. One way of approaching this problem might be to find the "best" n^{th} order filter by the following procedure. First, we would write the unknown filter F(s) in terms of 2(n-1) parameters as

$$F(s) = \prod_{i=1}^{n-1} \frac{S + \alpha_i}{S + \beta_i} \tag{5.59}$$

It is easily verified by Eq. (5.37) that this gives an H(s) with denominator of order n. Next we could vary the 2(n-1) parameters $\alpha_1, \alpha_2, \ldots \alpha_{n-1}$, β_1, $\beta_2, \ldots \beta_{n-1}$ so as to minimize the resulting loop error variance as computed by Eqs. (5.38) through (5.40). This procedure would, of course, be impossible to carry out for n reasonably large. An alternative approach is to assume more of the structure initially, leaving less parameters to be varied.

Adopting the latter procedure, we notice that $|1 - H(j\omega)|^2$ is equal to the difference between a constant and a rectangle:

$$|1 - H(j\omega)|^2 = 1 - \left(1 - \frac{1}{1 + \Lambda\delta^2}\right) f(\omega) \tag{5.60}$$

where

$$f(\omega) = \begin{cases} 1 & |\omega| < \alpha \\ 0 & |\omega| > \alpha \end{cases} \tag{5.61}$$

A familiar way of approximating the rectangular function is with Butterworth functions of the form

$$f(\omega) = \frac{1}{1 + \left(\frac{\omega}{\alpha}\right)^{2n}} \qquad n = 1, 2, \ldots \tag{5.62}$$

It turns out that we will achieve much better performance by replacing α by a parameter β which we will vary to optimize the loop. With this change, and substitution of Eq. (5.62) into Eq. (5.60) we have

$$|1 - H(j\omega)|^2 = \frac{(\frac{\omega}{\beta})^{2n} + \frac{1}{1 + \Lambda\delta^2}}{1 + (\frac{\omega}{\beta})^{2n}} \qquad (5.63)$$

This will approach the optimum given by Eq. (5.58) both for very large and very small ω, with the speed of convergence controlled by n. A particular advantage of our choice of Butterworth functions is that in calculating $H(j\omega)$ (and $F(j\omega)$) we can take advantage of the fact that the factors of terms of the form $1 + x^{2n}$ have been previously studied.*

Performance of these approximations has been calculated for n = 1, 2, 3 and 4 and is shown in Fig. 16. These results are for a threshold defined by $VAR[e_\ell(t)] = 1/10$.

Another approach is to calculate the total loop error (again assuming rectangular input spectrum) for the second order loop discussed in section 5.3. The loop parameters are then chosen to minimize the loop error variance. This gives the result shown with a dashed line in Fig. 16. The fact that these two "optimized" second order loops give essentially the same threshold suggests that we have roughly established the limit to performance obtainable with a second order loop. However, we can see from Fig. 16 that significant improvements can be achieved with higher-order loops. For instance, with a modulation index of 10 the fourth order loop gives 2 db of improvement over the second order loop and comes within $1^1/_2$ db of the optimum.

*See, for example, p. 252 of Seshu and Balabanian, Ref. 3.

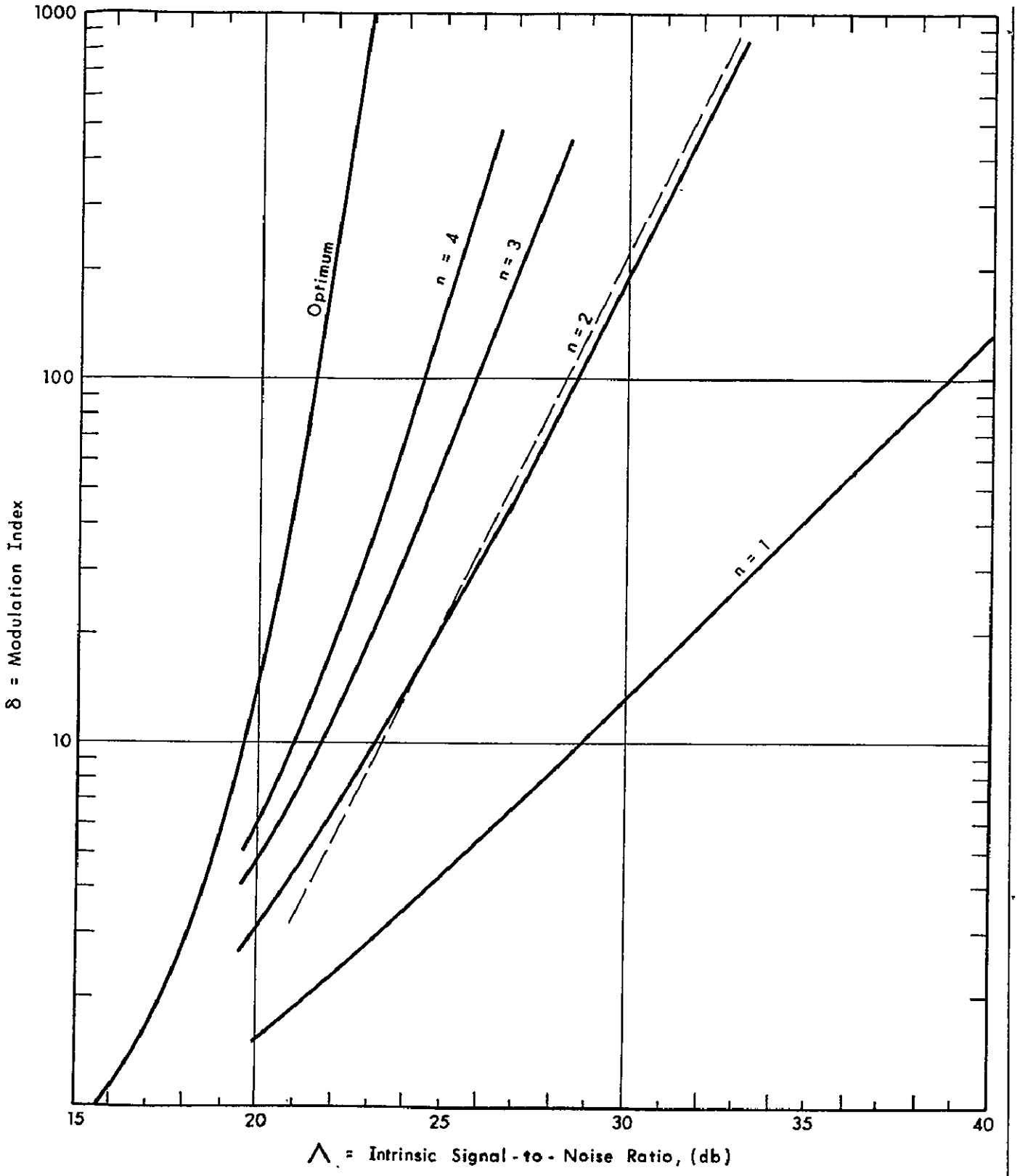


Fig. 16 Performance of approximations to the optimum transfer function.
(Var $[e_\ell(t)] = 0.1$)

REFERENCES - CHAPTER V

1. M. C. Yovits and J. L. Jackson, "Linear Filter Optimization with Game Theory Considerations," 1955 IRE National Convention Record, Part 4, pp. 193-199.

2. J. A. Develet, Jr., "A Threshold Criterion for Phase-Lock Demodulation," Proc. IEEE, pp. 349-356, February, 1963.

3. S. Seshu and Norman Balabanian, "Linear Network Analysis," Wiley, New York, 1959.

VI. TECHNIQUES FOR EXPERIMENTAL MEASUREMENTS

6.1 Previous ADCOM Work on Signal-to-Noise Ratio Measurement

Extensive laboratory study on the techniques of signal-to-noise ratio measurement has previously been performed at ADCOM. Results of this work have been reported in ADCOM reports.[1,2] The earliest of these measurement systems, shown in Fig. 17 provided useful data on the operation of an oscillating limiter at 455 kc.

Another system, shown in Fig. 18 was based on coherent separation of signal and noise components. The results obtained with this system were not so consistent as the previous system and the approach was dropped.[2]

Further work on signal-to-noise ratio measurement is reported in Ref. 1. The functional diagram of the system employed is shown in Fig. 19. This system measures the r-f signal-to-noise ratio and is usable after nonlinear predetection signal-with-noise processing. The present measurements do not require the use of this complex system since all linear circuits may be employed up to the detector circuits requiring measurement.

The system designed for the present project and currently under construction relies on the techniques which have proven effective in these previous studies. In particular, the first design mentioned above provides much of the basis for the present design.

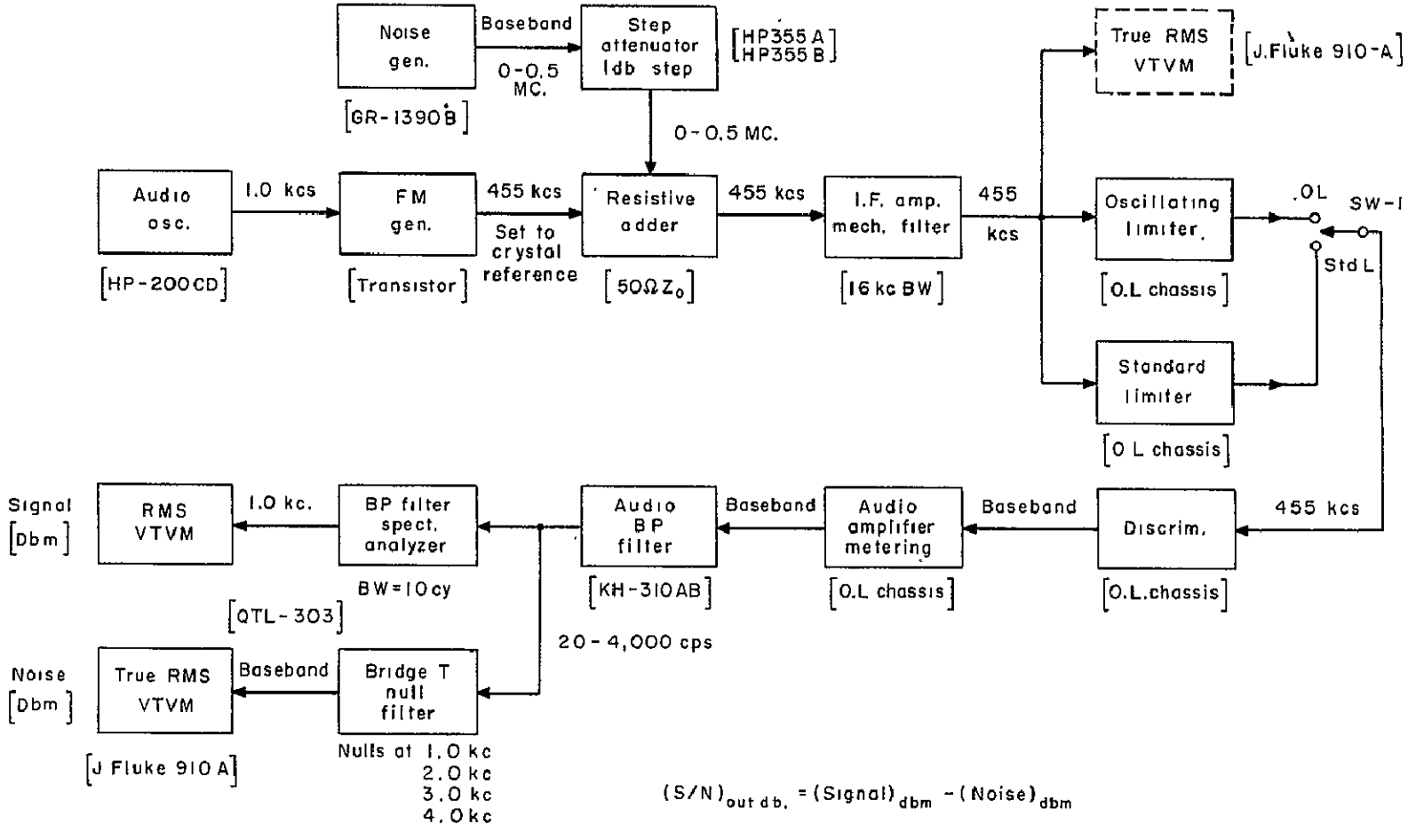


Fig. 17 Test system for the measurement of S/N ratio.

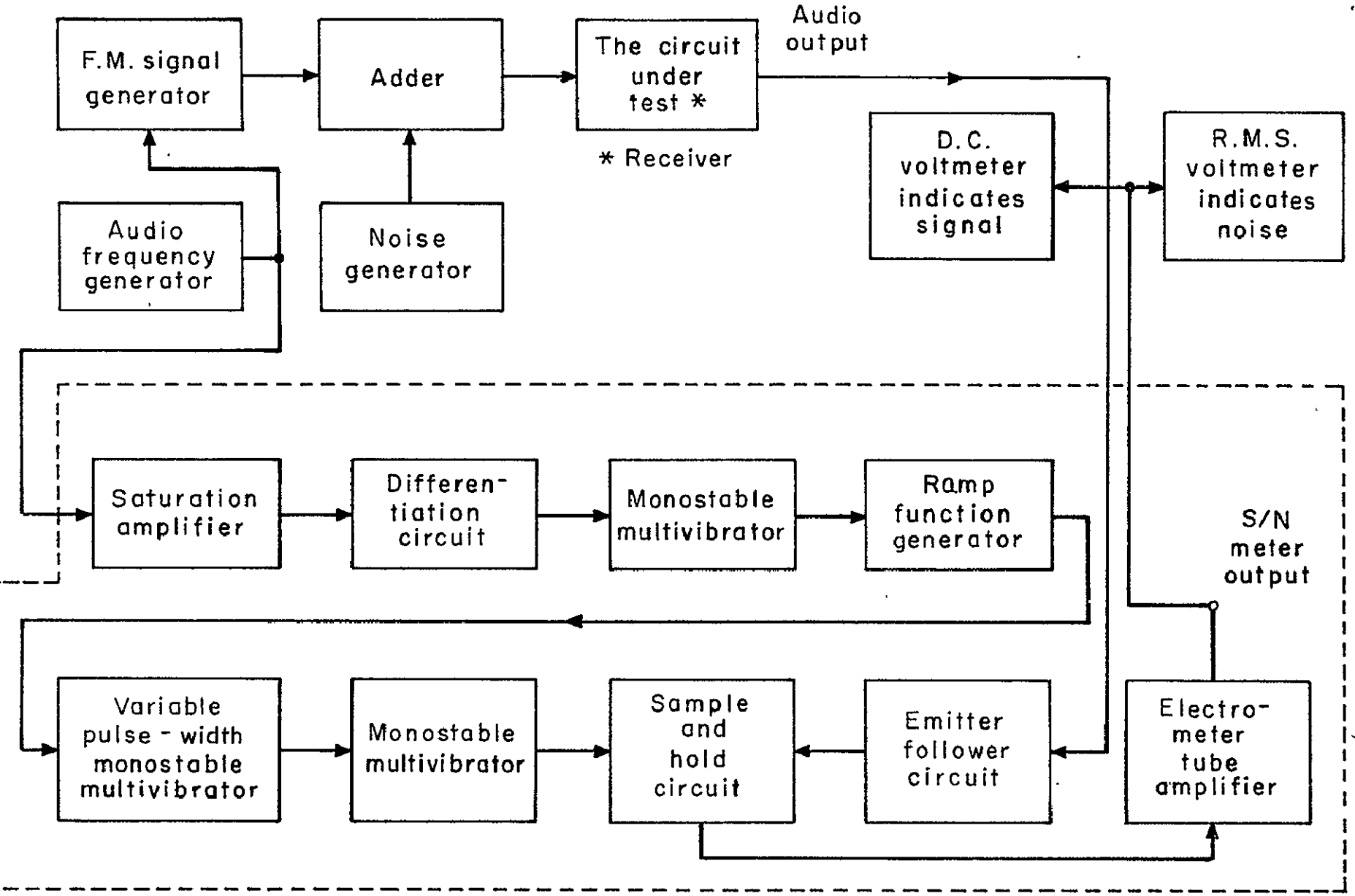


Fig. 18 S/N meter and block diagram of the experimental configuration.

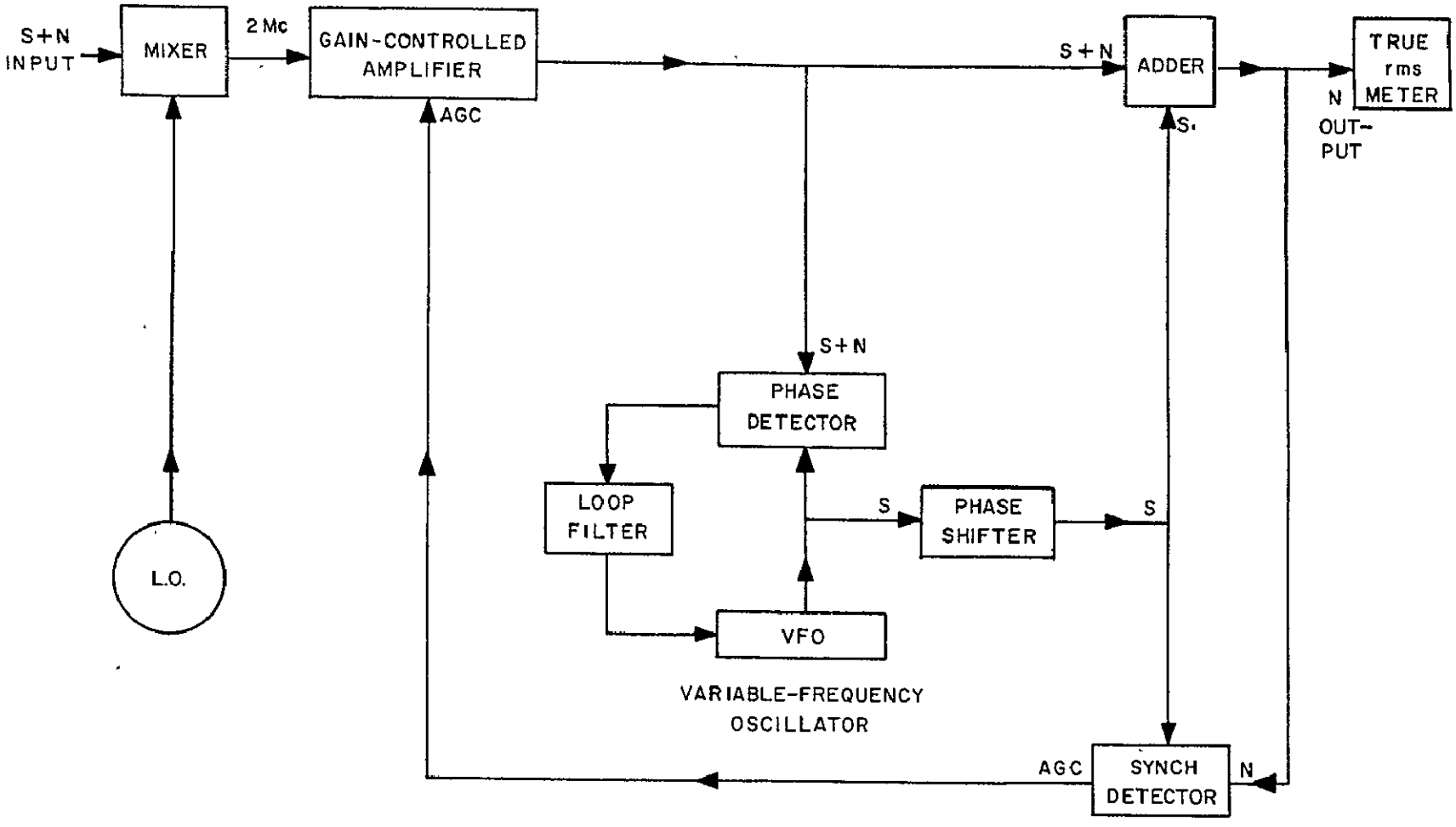


Fig. 19 Functional diagram for the ADCOM S/N meter.

6.2 Test System Design

The demodulator threshold test system is required to measure characteristics of the various FM demodulation schemes. The desire to obtain accurate results requires that the system allow for calibration repeatability to within one-half of a db (± 5%) overall. To achieve this accuracy requires that the short-term (15-30 minute) stability of many of the individual components be ± 0.1 db or better. The long-term stability need not be of high accuracy because provision is made to calibrate the system at intervals. However, the requirements for calibration should be held to a minimum by providing monitoring circuitry.

One method of providing calibrated signal and noise levels is shown in the diagram of Fig. 20. The FM signal is generated, passed through a precision calibrated attenuator, and added to a noise source that has also been passed through an attenuator. The composite signal plus noise is filtered in stable measurable filters to provide precise noise bandwidths.

The filtered composite signal is fed simultaneously to a wide bandwidth-limiter-discriminator (to be used as a standard of comparison for threshold improvements) and to the demodulator under test. The resultant demodulated signal is filtered with a calibrated lowpass filter. This provides a known baseband noise bandwidth for the precision measurements. The output signal level is measured with a narrowband spectrum analyzer and the noise level is measured with a true rms meter after the signal has been removed with a notch filter.

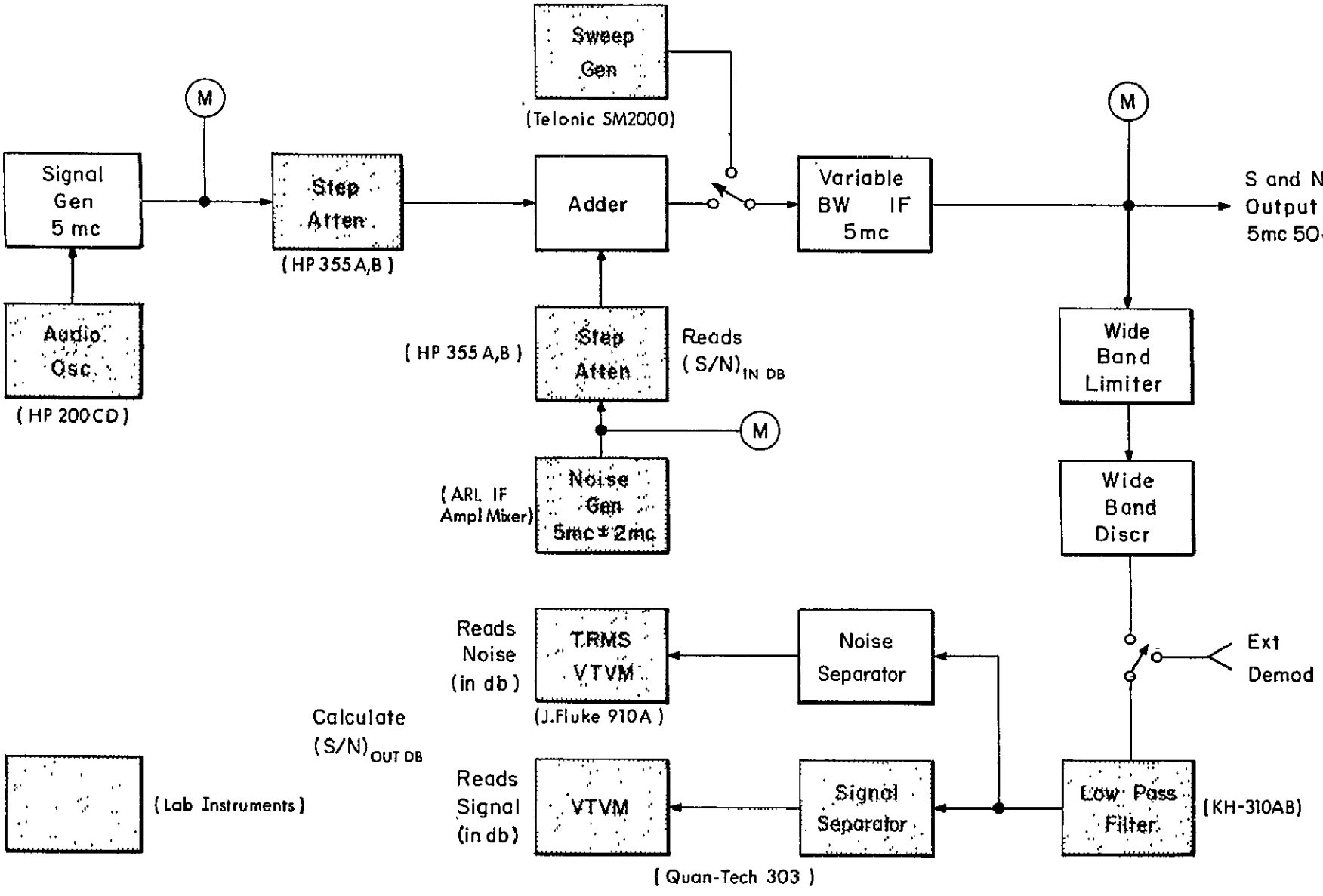


Fig. 20 Manual demodulator threshold test system.

Measurements on this system are taken by adjusting the input signal-to-noise ratio with the step attenuators in steps of one db. The output signal and noise levels are read and recorded. The threshold curve is plotted by calculating the signal-to-noise ratio at the output from the meter readings for each point on the curve.

6.2.1 Automatic Threshold Curve Plotter

The addition of certain laboratory equipment allows the time consuming point-by-point method to be replaced by a rapid and more accurate automatic plotting system. This system is shown in Fig. 21.

The basic system remains unchanged with two exceptions. The step attenuator in the noise channel is replaced by a motor driven attenuator and the signal portion of the demodulator output is held constant by an AGC circuit. Provision is made for applying a d-c voltage proportional to the input signal-to-noise ratio (in db) to the x input of an x-y plotter. The y input of the plotter is taken from the true rms noise measurement and passed through a d-c logarithmic amplifier so that it is proportional to the output signal-to-noise ratio in db.

The required additions to the system in order to provide the greatly increased accuracy and repeatability of the automatic system are

a) the motor driven attenuator, and
b) the AGC control circuit.

The other components of the system are standard commercial laboratory equipment.

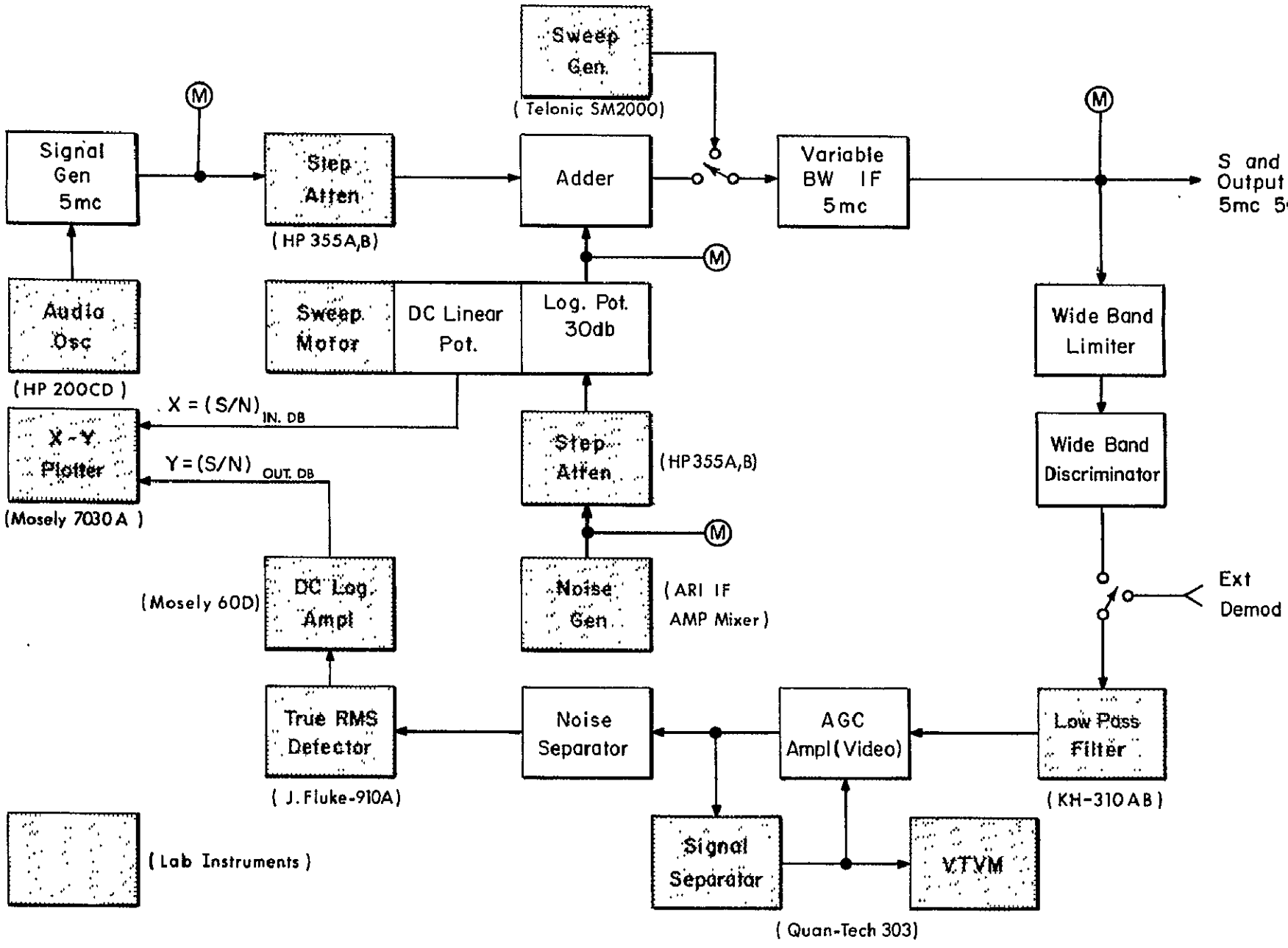


Fig. 21 Automatic demodulator threshold test system.

REFERENCES - CHAPTER VI

1. Baghdady, E. J., and Marshall, A. C., "FM Improvement Techniques," Final Report, Contract AF30(602)-2666, ADCOM, Inc.; Cambridge, Mass., February 1, 1963.

2. "Improving the Reliability of Telemetry Reception," Final Report, ADCOM, Inc., Contract NAS8-1688, January 31, 1962.

VII. OUTLINE OF WORK FOR THE NEXT INTERVAL

The work planned for the next interval may be divided into a theoretical part and an experimental part.

7.1 Outline of Theoretical Work

This part of the planned work, expected not to exceed one third of the effort, will be as follows:

a) Complete the evaluation of the work by Akima, Battail, Schilling and Billig, Van Trees and Viterbi.

b) Complete the analysis of the band-dividing technique based on sample-and-hold message representation to determine

   (i) Effects of filter mismatch and detuning, and the associated AM and FM transients giving rise to inter-sample interference and erroneous estimation of the frequency.

   (ii) Devise and refine schemes for minimizing effects in (i).

   (iii) Determine the effects of phase steps in going from one filter output to the other (the incoherency problem) and evaluate methods for avoiding these steps or abating their effects.

c) Complete the analysis of combined PLL and FCF demodulation technique.

d) Complete evaluation of other approaches specified in the Work Statement.

7.2 Outline of Experimental Work

This part of the planned work, expected to amount to a major fraction of the effort, will be as follows:

a) Complete laboratory testing preparations.

b) Complete design evaluation analyses of the new implementation concepts, developed on this program, of the band-dividing technique and related circuitry.

c) Complete design and implementation of a higher-order phase-lock demodulator for early preliminary tests, and for providing a test bed for other promising variations on the basic phase-lock technique.